Contribution à l'étude des retours haptiques pour améliorer l'expérience audiovisuelle

Fabien Danieau

ANNÉE 2014

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Informatique

École doctorale Matisse

présentée par

Fabien DANIEAU

préparée au centre Inria Rennes - Bretagne Atlantique
et à Technicolor R&I

Contribution to the Study of Haptic Feedback for Improving the Audio-Visual Experience

Thèse soutenue à Rennes
le 13 Février 2014

devant le jury composé de :

Pr. Sriram SUBRAMANIAN
Professeur à l'Université de Bristol / *Président*

Pr. Géry CASIEZ
Professeur à l'Université de Lille 1 / *Rapporteur*

Pr. Abdulmotaleb EL SADDIK
Professeur à l'Université d'Ottawa / *Rapporteur*

Pr. Alain BERTHOZ
Professeur au Collège de France / *Examinateur*

Dr. Anatole LÉCUYER
Directeur de Recherches chez Inria / *Directeur de thèse*

Dr. Marc CHRISTIE
Maître de Conférences à l'Université de Rennes 1 / *Examinateur*

Dr. Philippe GUILLOTEL
Chercheur Senior à Technicolor / *Examinateur*

Dr. Julien FLEUREAU
Chercheur Senior à Technicolor / *Examinateur*

À Fanny

Remerciements

Tout d'abord je voudrais remercier Anatole Lécuyer et Philippe Guillotel qui ont initié ce projet de thèse et qui m'ont encadré pendant ces trois ans. J'ai pris beaucoup de plaisir à travailler avec vous et j'ai énormément appris à vos côtés. Comme vous me l'aviez dit au début, ces trois ans sont passés très vite. Ce fut une expérience inoubliable, j'en garderai un excellent souvenir.

Je remercie vivement Julien Fleureau, Nicolas Mollet et Marc Christie qui ont aussi activement encadré cette thèse. Vos conseils et vos encouragements durant ces trois années ont été précieux et ont concrètement permis l'aboutissement de ce travail.

Je voudrais également remercier les membres du jury qui ont accepté du juger mon travail et mon manuscrit. Je remercie ainsi Pr. Géry Casiez et Pr. Abdulmotaleb El Saddik d'avoir accepté d'être rapporteurs de mon manuscrit, et Pr. Alain Berthoz et Pr. Sriram Subramaniam d'avoir accepté d'être examinateurs. Merci également pour vos commentaires pertinents sur mon travail et pour vos conseils.

Merci à toute l'équipe Hybrid du centre Inria/IRISA et à tous les membres des anciennes équipes VR4I et Bunraku. Merci de m'avoir accueilli, nous avons passé de super moments ensemble, aussi bien au laboratoire qu'en soirée. Merci à Orianne Siret, Emmanuel Badier, Steve Tonneau, Ferran Argelaguet Sanz, Julian Joseph, Florian Nouviale, Jonathan Perrinet, Loiez Glondu, Gabriel Cirio, Thomas Lopez, Kévin Jordao, Jonathan Mercier-Ganady, Hui-Yin Wu, Teofilo Dutra, Carl-Johan Jorgensen, Rozen Bouville-Berthelot, Jérôme Ardouin, Huyen Nguyen, Anne-Hélène Olivier et Julien Bruneau. Un merci particulier à mes deux collègues de bureau Anthony Talvas et Merwan Achibet. Merci à Nathalie Denis pour sa gentillesse et son aide.

Merci à toute l'équipe de Technicolor R&I. J'y ai également reçu un accueil chaleureux et passé de très bons moments. Là encore il y a beaucoup de personnes à remercier et je ne pourrai pas être exhaustif. Merci en particulier à Mehmet Türkan, Martin Alain, Ronan Boitard, Matis Hudon et Mikaël Le Pendu pour tout ce que l'on a partagé ensemble. Merci à l'équipe exploratoire et aussi à l'équipe HMIUB qui m'a adopté sur la fin de ma thèse, et où je suis heureux de rester pour continuer mes recherches. Je voudrais également remercier les stagiaires qui ont travaillé avec moi sur ce projet de thèse et qui m'ont grandement aidé. Merci à Audrey Cabec, Jérémie Bernon et Inès Rouxel.

J'aimerais remercier l'équipe Flowers du centre Inria Bordeaux où j'ai fait mon stage de fin d'études d'ingénieur. C'est là que j'ai découvert et fais mes premiers pas en recherche. Merci à Pierre-Yves Oudeyer et Pierre Rouanet de m'avoir accueilli, formé et poussé à continuer. Un grand merci à Pierre et Matthieu Lapeyre d'avoir fait mille kilomètres pour venir me voir.

Merci à tous mes amis que j'ai connus et laissés à Bordeaux. Vous me manquez beaucoup et c'est toujours un plaisir de venir vous voir. Votre soutien et vos encouragements m'ont énormément aidé. Merci à François Tessier, Marion Gimenez, Patrick Dos-Santos, Regis Cobrun et François Bégou.

Merci à mes nouveaux amis bretons, de souche ou de cœur. C'est avec joie que je reste en Bretagne avec vous où je suis sûr que nous continuerons à passer de bons moments. Merci à Fabien Racapé, Darya Khaustova, Anna Mateus, Simon Malinowski, Marco Bellaviqua et Youssef Alj.

Je voudrais terminer en remerciant ma famille pour leur soutien depuis le début de mes études, mes parents Gilles et Sylvie, ma sœur Aurélie et ma grand-mère Monique. Enfin, merci à ma Fanny qui me soutient quotidiennement et grâce à qui j'en suis là aujourd'hui.

Contents

Chapter 1

Introduction

The sense of touch (haptics) in interactive systems has been particularly studied and appears to be a key factor in user immersion [Rei04]. Many haptic interfaces that enable the physical interaction with virtual or remote objects have been developed and evaluated in the past decades [BS02, HACH+04]. They were first used in virtual reality or teleoperation systems. Nowadays haptic technologies are used in various applications in medical, robotics and artistic settings.

In contrast, the use of haptics in a multimedia context, in which haptic feedback is combined with one or more media such as audio, video and text, seems surprisingly underexploited. Yet, in 1962, Heilig introduced the *Sensorama*, a system where one could watch a 3D movie, sense vibrations, feel the wind and smell odors [Hei62]. Although the potential industrial impact of haptics for audiovisual entertainment seems to be important, research and technology remains essentially focused on improving image and sound. Only a few systems, known as "4D cinemas", currently exploit this technology. However, the number of articles reporting the potential of haptic feedback for multimedia is increasing. O'Modhrain et al. have demonstrated that the benefits of haptic feedback observed in virtual reality, video games or telepresence are applicable to multimedia applications [OO03]. Haptic feedback increases user's feeling of presence, of realism and user's engagement in the application. Haptic feedback may also open new ways to experience audiovisual content. The relation between users and audiovisual content is no longer limited to a passive experience but could enable physical involvement in a more immersive experience [MTB06]. More than physical sensations associated to audiovisual content, the user could expect to receive a complementary piece of information or to intensify an emotion through haptic interaction. The combination of haptics and audiovisual content becomes the complete medium of haptic-audiovisual (HAV [EOEC11]) content. Worthy of study, HAV is thus becoming a new scientific field with its own specific requirements and challenges.

The young field of study of HAV introduces new scientific issues. How can haptics be employed efficiently in conjunction with image and sound, and how can haptic feedback be created for this purpose? What kind of device is suitable for rendering haptic feedback in a viewing scenario (cinema or user living space, potentially shared)?

Moreover, to what extent can haptics influence the video viewing experience? and how can the resulting quality of experience be evaluated? These issues are organized and studied in this manuscript.

1.1 Adding haptic feedback to audiovisual content: challenges and workflow

The issues of adding haptic effects to audiovisual content can be organized within a workflow illustrated in Figure 1.1. Inspired from the typical workflow for video-streaming [WHZ+01], it comprises three stages: production, distribution and rendering. We use the term "haptic effect" to designate the use of a haptic feedback in audiovisual content, a generalization of the term employed in the specific context of video viewing [OO03, YAMS06, CES09].

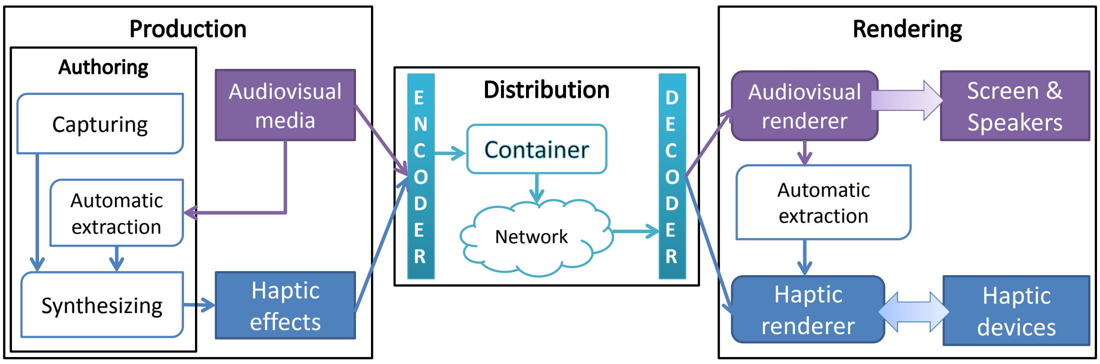


Figure 1.1: Workflow for adding haptic effects to audiovisual content. In this thesis, we consider haptic effects as a component of a multimedia content. Effects are typically: (i) produced, (ii) distributed and (iii) rendered in the user living space in parallel to the audiovisual content.

The first stage in the workflow deals with the **production** of the content, i.e. how haptic effects can be created or generated in synchronization with the audiovisual content. Three techniques emerge from the literature: the capture and processing of data acquired from sensors, automatic extraction from a component of the audiovisual content (image, audio or annotations) and manual authoring of haptic effects.

The second stage in the workflow deals with the **distribution** of haptic effects. Current technologies allow mass distribution of media over networks, so there is a strong requirement for haptic effects also to be distributable in this way. This raises questions on formalizing haptic effects. The synchronized transmission of haptic effects over networks is termed haptic broadcasting [CHK+09].

Finally, in the third stage, an encoded haptic effect is **rendered** on a specific haptic device and experienced by the user, while the audiovisual content is displayed on a screen and played on speakers. Haptic effects are converted into commands for the haptic device by the dedicated haptic renderer.

A last main aspect, complementary to the workflow, is the evaluation of the user experience which cuts across production, distribution and rendering. The quality of experience (QoE) has several definitions [Jai04, Kil08] but can be defined in our context as the measure of the user' subjective experience with an audiovisual content. Most interest to date has focused on the technical aspects of the three stages of the workflow, but there is also a clear necessity to measure the quality of haptic-enhanced audiovisual experiences.

1.2 Thesis objectives

The work presented in this manuscript belongs to the recent field of study of haptic-audiovisuals (HAV). Numerous improvements of the three stages of the workflow (production, distribution and rendering of haptic effects) are required in order to make HAV a mature technology. At the time of starting this thesis, the aspect of formalization and transmission of haptic effects were partly studied and on the way to be standardized by the MPEG[1] group. This stage will not be considered in this work. This thesis focuses therefore on the two stages of production and rendering of haptic effects. Four points will be addressed in particular (see Figure 1.2): (1) designing a new **device** dedicated to audiovisual viewing, (2) improving the **haptic rendering** for haptic-audiovisual scenarios, (3) simplifying the creation of haptic effects thanks to a new **authoring tool**, and (4) enriching the taxonomy of haptic effects with dedicated **cinematographic effects**. These four objectives are detailed in the following sections.

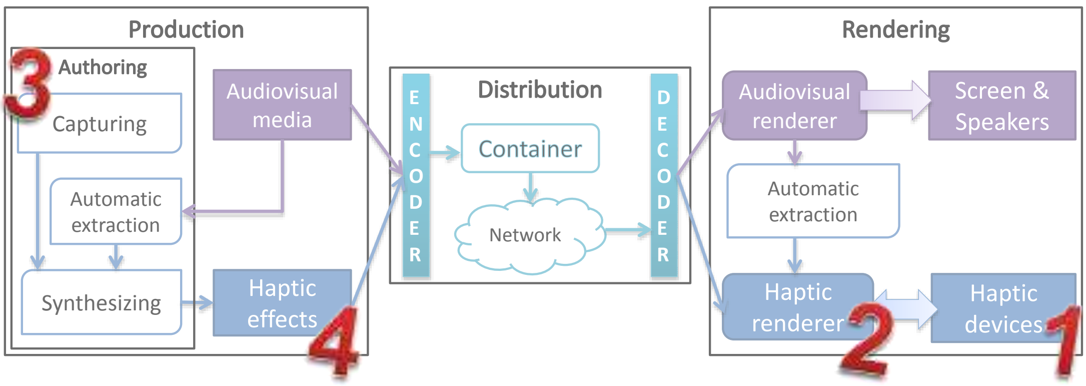


Figure 1.2: Overview of the four objectives and main contributions of the thesis.

1.2.1 New haptic device dedicated to video viewing settings

Numerous haptic interfaces have been designed for virtual reality, teleoperation or gaming purposes [BS02, HACH$^+$04], but few exist to enhance a video viewing session.

[1]ISO/IEC Moving Picture Experts Group

Moving seats, designed for amusement parks or "4D cinemas", provide strong effects of motions, vibrations or water sprays. But these cumbersome devices are not really adapted to traditional movie theaters or a user's living space. More popular devices are those embedding vibrating motors, such as gamepads or mobile phones. But the range of sensations they provide is quite limited. Therefore a real need for devices providing a wide range of sensations while remaining simple, comfortable, robust and adapted for a classical video viewing experience appears.

1.2.2 New haptic rendering algorithm for haptic-audiovisual scenarios

In typical haptic rendering algorithms, the haptic feedback is computed from the interaction between the user and the simulation [SCB04]. When the user touches a virtual object, the reaction force is computed and rendered on the haptic device. The haptic rendering is performed continuously during the simulation. In a context of haptic-audiovisual applications, haptic effects may have been designed by content creators to be triggered at specific moments in the audiovisual content (synchronized to sounds effects, movement of a character, explosions, etc.) [WRTH13, Kim13]. Such effects are designed independently from a specific device (and its constraints) and may occur in a noncontinuous way. Classical haptic rendering algorithms are not suitable in this context. New algorithms need to be developed to handle the transitions between effects and to adapt the haptic feedback to the workspace of the device.

1.2.3 New authoring tool for creating of haptic effects

The creation of haptic effects is often a manual process achieved by providers of hardware for "4D cinemas" [DBO, MED]. Effects are added on top of existing audiovisual content. These effects are not designed by the creators of audiovisual content and thus not fully integrated as part of the media. But the authoring of haptic effects is a complex task and few editors exist. For example, creating a 6DoF[2] effect of motion requires the edition of multiple parameters (usually three translational accelerations and three angular speeds). Therefore user-friendly tools are needed to design and add haptic effects to videos, similar to those used by movie-markers to design sound or visual effects during the (post-)production of audiovisual content.

1.2.4 New haptic effects for enriching the haptic-audiovisual experience

In virtual reality or teleoperation applications, haptic feedback allows the users to touch what they see. The first approaches combining haptics and audiovisual follow this principle (although there is no actual interaction with a video) [OO04, LLC+05, GMS06]. But an audiovisual content may represent more than physical events. Emotion or information can also be conveyed (with romantic movies or documentaries for instance). Then the traditional use of haptic feedback would be quite limiting while haptics has

[2] Degrees of Freedom

the potential to convey more than physical sensations [MTB06]. New ways of combining haptics and audiovisual have to be studied in order to enrich the taxonomy of haptic effects.

1.3 Approach and contributions

This manuscript describes the work we carried out in order to address the four objectives: designing a new device dedicated to video viewing settings, improving the haptic rendering for haptic-audiovisual scenarios, simplifying the creation of haptic effects thanks to a new authoring tool, and enriching the taxonomy of haptic effects with dedicated cinematographic effects.

We first present in **Chapter 2 the related work on haptic-audiovisuals**. The key challenges of the field are detailed, and previous works for producing, distributing and rendering haptic effects are reviewed. Then the current techniques and metrics to evaluate the user haptic-audiovisual experience are described.

The following chapters are dedicated to the scientific contributions we proposed. They are gathered in two parts: **Part I** describes contributions related to the **rendering** of haptic effects and **Part II** details contributions related to the **production** of haptic effects. We have used a user-centered approach and we have designed metrics to measure the quality of experience. Our contributions were systematically evaluated in this thesis.

1.3.1 Part I - Rendering haptic effects: novel device and algorithms for rendering haptic effects in video viewing settings

The rendering of haptic effects is a key challenge to enhance the video viewing experience. In this last stage of the HAV workflow, haptic effects are delivered to the end-user while the audiovisual content is displayed. The haptic feedback provided depends thus on the capabilities of the haptic device used. As explained in Section 1.2.1, there is a lack of devices rendering rich haptic feedback in consumer environment.

In **Chapter 3 we propose a novel device to render haptic effects in a video viewing scenario**. More particularly we focus on the rendering of the sensation of motion. Motion simulators moving the whole user's body are quite expensive and cumbersome for a classical video viewing experience. Hence we introduce a new device, the *HapSeat*, in which the sensation of motion is provided by the stimulation of three points of the user's body (head and hands). A proof-of-concept has been designed and uses three low-cost force-feedback devices. Two control models have been implemented. A user study has been conducted to evaluate the relevance of this concept and the impact of the different models on the quality of experience.

Haptic effects can be created independently of a haptic device and may occur in a noncontinuous way during the display of the audiovisual content. The haptic rendering needs to adapt the haptic effects to the workspace of the device used, and has to handle the transitions between effects. Classical haptic rendering algorithms are not suitable in this context.

In **Chapter 4 we propose a new haptic rendering algorithm for haptic-audiovisuals**. We introduce the use of washout filters for force-feedback devices. The principle of a washout filter is to move a device toward the center of the workspace under the user's perception threshold. This technique is well-known for the control of traditional motion simulators. As the whole user's body is moving, the perception threshold is determined by the capabilities of the human vestibular system. This approach is however not applicable for the force-feedback devices, including the *HapSeat*, which stimulates the human kinesthetic system. We thus propose to rely on a biomechanical model to compute the user's perception thresholds. Based on this technique, we optimize the haptic rendering, allowing the generation of multiple effects of motion. A user study was also conducted to quantify the benefits of this washout filter on the quality of experience. Evaluations were performed in order to characterize the user's perception of this washout filter, and a study on an actual haptic-audiovisual movie was conducted to generalize these results.

1.3.2 Part II - Producing haptic effects: tools and techniques for creating haptic-audiovisual content

The second part of the manuscript is focused on the production of haptic effects. Our approach is to consider haptics as a new medium, equivalent to image and sound. Therefore we propose several techniques to design and associate haptic effects to audiovisual content.

In **Chapter 5 we present new methods for editing haptic and motion effects**. The design of motion effects is particularly challenging due to the 6DoF which all have to be set at a time. Thus we propose three different methods: two manuals methods using a force-feedback device as input and one automatic method to capture motion effects during the recording of a video sequence. These methods were implemented in a new authoring tool: the *H-Studio*. Besides we introduce a new feature to preview motion effects thanks to a force-feedback device. This way the haptic designer can preview the amplitude of the effects, the dynamic of the sequence and the synchronization of the effects with the video. Eventually a user study has been conducted to evaluate the preview of the captured motion effects.

Few haptic-audiovisual systems have been described in the literature, but it appears that haptic effects are mainly used to make the audience feels physical events happening in a video. This approach is similar to the use of haptic feedback in virtual reality applications. However we believe that haptics could be combined to other aspects of an audiovisual content in a same way that sound in movies can be related to physical events (sound effects) or to the ambiance (music).

In line with this observation, we introduce in **Chapter 6 the concept of *Haptic Cinematography*** which presents haptics as a new dimension in the creation space for filmmakers and we propose a taxonomy of haptic effects. We detail in particular new haptic effects coupled with classical cinematographic camera motions to enhance video viewing experience. More precisely we propose two models to render haptic effects based on camera motions. The first model makes the audience feel the motion of the camera and the second provides haptic metaphors related to the semantics of the camera effect. A user study has been conducted to evaluate the impact of these haptic effects on the quality of experience.

Finally **Chapter 7 provides conclusions and perspectives** of the work presented in this manuscript.

Chapter 2

Related work on haptic-audiovisuals

Contents

Haptic technology has been widely employed in applications ranging from teleoperation and medical simulation to art and design, including entertainment, flight simulation, and virtual reality. Today there is a growing interest among researchers in integrating haptic feedback into audiovisual systems. A new medium emerges from this effort: haptic-audiovisual (HAV) content.

The aim of this chapter is to survey the results obtained in this young field of research and identify its key challenges. We build on the workflow for adding haptic effects to visual audiovisual content to detail contributions related to its three main stages (see Figure 2.1). Firstly, haptic effects are produced thanks to dedicated tools and techniques, reviewed in Section 2.1. Secondly, haptic effects are formalized and combined to audiovisual contents in order to be distributed to end-users, potentially through streaming platforms. Various models and formalizations for the distribution are presented in Section 2.2. Thirdly, haptic effects are rendered by dedicated haptic interfaces. Section 2.3 offers an overview of the wide range of published renderers classified by the type of device (wearable, handheld, desktop or seat). We finally discuss techniques for evaluating the quality of experience (QoE) with such systems in Section 2.4. We conclude by discussing the developing prospects in the field.

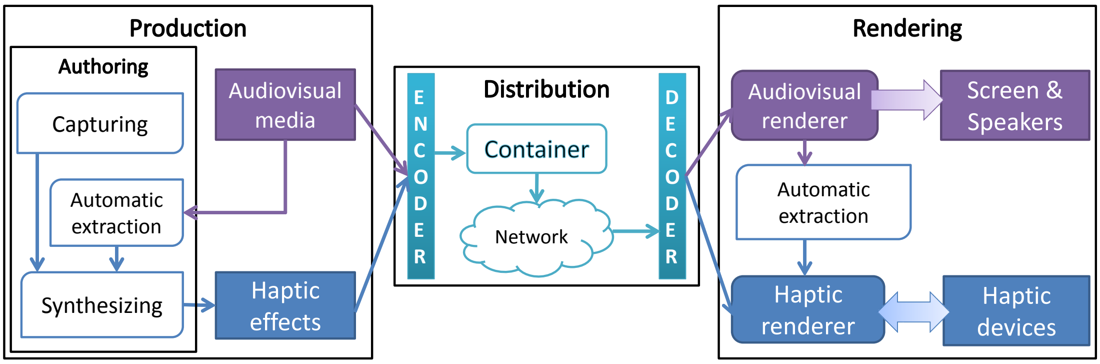


Figure 2.1: HAV workflow for adding haptic effects to audiovisual content.

2.1 Production of haptic effects

Production is the task of creating haptic effects in order to enhance audiovisual content. Three methods to create them have been reported in the literature: (i) capturing haptic effects from the real world using physical sensors, (ii) generating haptic effects by an automated analysis of audio and/or visual contents, and (iii) manually synthesizing haptic effects from scratch or by editing effects obtained with the previous methods. Haptic effects will be classified according to their perceptual characteristics (tactile, kinesthetic, and proprioception).

2.1.1 Haptic effects for audiovisual content

The MPEG-V format is a promising standard allowing to describe sensory effects for multimedia content (see Section 2.2.1.1) [MPE11]. A classification of sensory effects such as taste, smell and haptic is proposed. Haptic effects reported were temperature, wind, whole body vibration, water sprayer, passive kinesthetic motion and force (the user simply holds a force-feedback device), active kinesthetic (the user can explore actively the content thanks to a force-feedback device), tactile and rigid body motion (the whole body of the user is moved as in motion simulators). In the standard the effects are defined in an abstract way to capture the haptic designer's intention. The description is supposed to be device independent but some effects are directly linked to a specific device (water sprayer or tactile for instance).

In contrast, the classification we propose is based on haptic perceptual capabilities. Haptic feedback is often separated into two categories: tactile and kinesthetic feedback. There are three types of tactile stimuli: perception of vibration, of pressure [Shi93] and of temperature [JB02]. Two types of kinesthetic stimuli may be defined [Jon00]: perception of movement and limb position, and the perception of forces. Finally, haptic perception may result from the motion of the user's own body [Ber00]. Both the vestibular system and proprioception contribute to the perception.

We then propose a table summarizing haptic effects in HAV systems in which each category is mapped to contributions from the literature (see Table 2.1). The reader may also refer to the guidelines for the design of vibrotactile effects [vE02] or haptic feedback in multimodal environments [HS04]. These individual effects can be combined to create more complex effects. For example, the haptic effect associated with an explosion might be defined with a combination of temperature and vibration.

Haptic effects are mostly used to represent physical events which occur in the scene (see references in Table 2.1). The user perceives stimuli which are directly related to the audiovisual content (e.g. bumps when driving off-road), augmenting the physical event and the sense of "being physically present". However other aspects of an audiovisual content, such as ambiance, can be enhanced as well [KCRO10]. The role of haptic effects in audiovisual content is analogous to that of audio in movies: audio is used for increasing the realism (sound effects) and to create ambiance (music). In movies, a clear separation is drawn between diegetic sounds (a sound for which the source belongs to the diegesis, the recounted story) and non-diegetic sounds (a sound for which the source is neither visible nor implied in the action, typically such as a narrator's comment or mood music). Non-diegetic haptic effects have similar potential. Non-visual content could be augmented by providing additional information that is perceived by the user.

The use of haptic effects to enhance ambiance or emotion is not straightforward. The haptic effect designer may explore results from research on affective haptics: recent works attempt to communicate affect with haptic feedback [SM07] or trigger users' emotions with the help of haptic patterns [TNP+09, LCB+09].

2.1.2 Capturing haptic effects from the real world

One approach for creating haptic effects is to capture haptic data related to an object or actor in a scene. The capture is performed by a "haptic camera" which is a physical sensor extracting haptic properties from the real world [Mac96]. Piezo-electric sensors can be used to capture forces [OO03] or vibrations but, most of the time, accelerometers are used to record accelerations and deduce forces applied to the targeted object. Brady et al. equipped a radio-controlled car to capture accelerations on X, Y and Z axes [BMO+02]. These recorded data were then directly transmitted and rendered to the user's control device. Recorded accelerations on X and Y axes control an embedded

Haptic Perception	Haptic Effect	Reference
Tactile	Temperature	[Dio97, PWN+]
	Vibration	[LLC+05, uR08, LCB+09, WRTH12, RAC10, KCRO10, PWN+]
	Pressure	[SKI05, PWN+, HTIS10, AMS11]
Kinesthetic	Movement	[GMS06, KPK+11]
	Force	[OO04, YAMS06, CES09, KPK+11]
Proprioception	Body Motion	[DBO, CJ4]

Table 2.1: List of potential haptic effects for audiovisual content. Individual effects can be combined to create complex effects.

2DoF force-feedback device and acceleration on the Z-axis drives a vibration device. In a less direct way, Kuchenbecker et al. recorded haptic events in a database to enable replay later [KFN05]. The authors recorded accelerations resulting from the impact of a stylus on different materials (wood, foam). These accelerations were transduced into forces and replayed by a force-feedback device when the user touched virtual materials.

A second approach consists of capturing haptic effects related to a whole scene. Depth (or 2.5D) cameras have been used to build touchable images [CES09, RS11]. A more precise result could be obtained with 3D trackers [MKT+05] but these devices are more expensive and the analysis of the scene would take longer (see Figure 2.2). The problem of capturing haptic effects remains strongly constrained by the available hardware. In contrast to video and sound recording, only a limited number of devices exist, mainly accelerometers and 3D cameras with considerable variations in precision and cost.

2.1.3 Automatic extraction of haptic effects from audiovisual content

Haptic effects can also be created automatically by extraction. The key idea is to generate haptic effects which are consistent with media content in order to highlight specific aspects. For example a scene showing an explosion could be enhanced by haptic feedback such as vibrations and heat. Video and sound analysis might be used to detect explosions and then automatically add haptic effects.

Automatic extraction can occur in the production stage or in the rendering stage. In the production stage, haptic effects are automatically generated and can be modified by the creator. In the rendering stage, haptic effects are automatically generated on the client side.

2.1.3.1 Generation from visual content

A classical way to extract content from an audiovisual media consists in using video analysis techniques. Typical algorithms rely on feature detectors to extract points of interest inside an image to build derived information (e.g. object identification) [TM07]. There are significant variations in the features they offer such as robustness to light

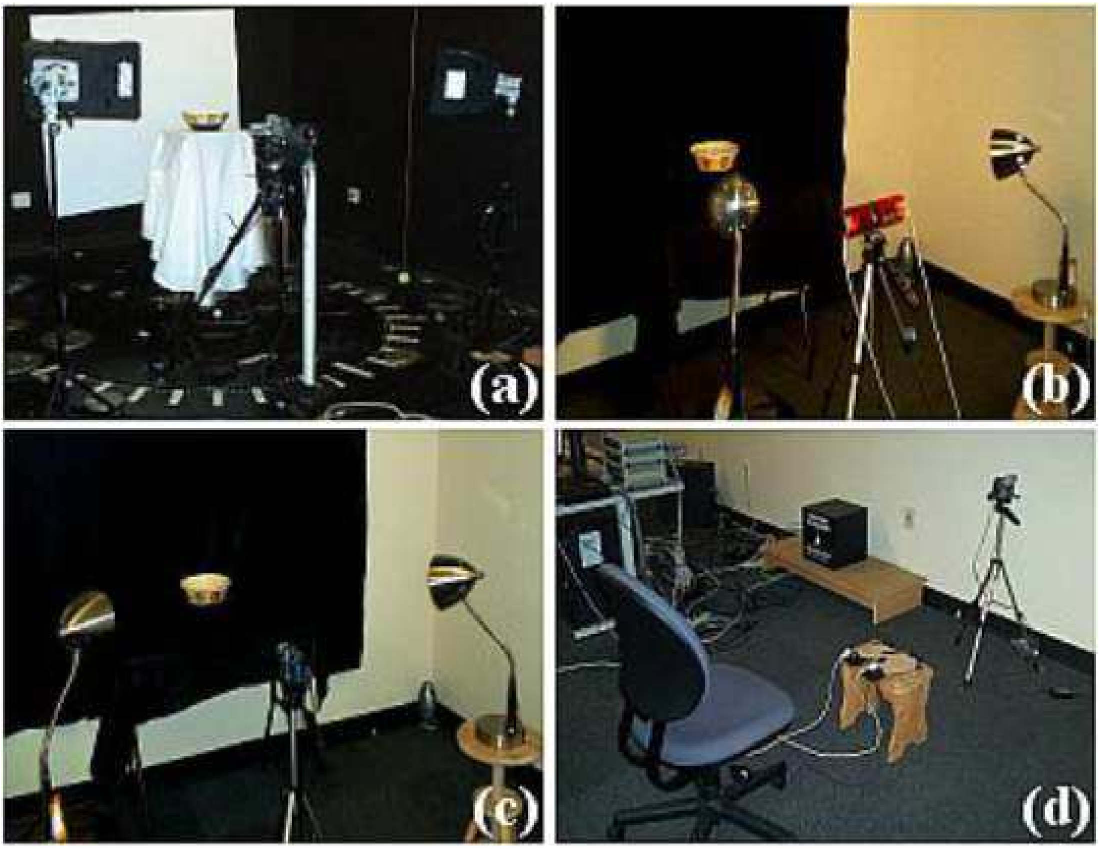


Figure 2.2: Capture of visual and haptic cues [MKT+05]. 360° images (a), shape and size (b), information about material and texture (c), and data from the haptic manipulation (d) are captured.

variations or motion, and computational cost. Some specific algorithms are dedicated to the detection of specific features such as faces [ZCPR03] or motion [HTWM04]. Detecting events is also possible. Video abstraction [TV07] and video data mining [ZWE+05] have both been used for event detection but are restricted to specific subjects such as sports, where the potential range of events is limited. Once a targeted event is detected in the audiovisual content, a haptic effect could be generated. For instance, Réhman et al. have shown how to automatically extract events from a soccer game video and to display them with a vibrotactile device [uR08]. Five vibration patterns were designed to represent the position of the ball on the field, to the team leading the game or to the goals. However the main focus was on how to render the effects rather than the video analysis. Rasool and Sourin have described several techniques to extract haptic properties from images [RS11]: haptic geometry from stereoscopic images, haptic textures based on shading information or image segmentation into haptic regions with constant physical properties. Kim et al. have relied on a saliency map to drive a vibrotactile array [KLC12]. A saliency map spatiotemporally abstracts perceptual importance in a visual scene. They have computed a map for each frame of a video and they have mapped the result to the vibrotactile array. Then vibrations are related to the most important elements of the video.

The difficulty of direct extraction of haptic information from video was pointed out by Mc Daniel et al. [MKT+05]. To simplify the problem, the authors built a database which maps visual information (a picture of an object) to haptic information (the 3D shape of the object). The database is used to generate appropriate haptic feedback for each object identified from visual information.

Even if computer vision provides a broad range of tools, most techniques to analyze and generate haptic feedback have not been yet explored in detail. The robustness and adaptability of the detection algorithms remain typical issues in the field [TM07].

2.1.3.2 Generation from audio content

Haptic effects can also be created from the audio content within audiovisual media. The main approach is to transduce an audible signal into a signal suitable for vibration motors. Chang and O'Sullivan used a band-pass filter to isolate frequencies compatible with a targeted vibration motor and then amplify and render the output signal on this device [CO05]. This system was developed for mobile phones which then vibrate according to ringtones. The "Integrator" development platform from Immersion is a similar commercially available system [IMM]. The "Reverb" module allows the automatical addition of haptic effects to any application using the output audio stream. The approach selected by Nanayakkara et al. is even more direct and does not require any processing of the audio stream [NTWO09]. The authors developed a chair for deaf people which renders music and vibration. The sound is played by speakers attached to the seat, which are specially designed to propagate vibrations to the surface they are attached to.

Most research follows this straightforward technique of the transduction of audio into vibrations. An alternative is to render specific parts of the audio signal. Chi et al. have developed an algorithm that detects pre-learned target sounds (gun effects) into a video game [CCO+08]. Once an effect is detected, a vibration pattern is render on the player's device. Lee at al. have developed a model which extracts perceptual variables of the audio signal, roughness and loudness, and converts them into a vibrotactile signal [LC13]. The model is able to detect specific sound effects such as explosions or the noise of a motor engine. The approach could be extended by attempting to represent the information conveyed by the audio stream. Audio analysis techniques to extract specific features would then be useful. For example the system described by Zhang and Kuo permits the identification of music, speech and environmental sound in an audio signal [ZK01].

2.1.3.3 Generation from metadata

Metadata can contain information about movements or physical properties of objects within the media. Yamaguchi et al. extracted data from a Flash [FLA] animation to compute force feedback as the user explores the content [YAMS06]. Since this format allows access to the geometry and position of elements within a 2D animation, it is possible to compute a force-feedback related to one of the objects in the scene. The authors defined a virtual mass for the targeted object and then computed a force-feedback relative to the acceleration and mass of this object. This technique can be applied to computer animations where a 3D model of the scene is available. But the system remains specific to animations and is not suitable for standard video. However some data formats allow for the description of audiovisual content. The MPEG-7 standard focuses on the description of multimedia content and can contain a description of movement within a scene [CSP02], opening many possibilities for the generation of haptic effects.

2.1.4 Graphical creation tools for synthesizing haptic effects

Although haptic effects can be created automatically, the need to create them before their integration with audiovisual content remains. Original effects may need to be edited. Neither of these functions can be automated.

Two main categories of graphical creation tools have been designed. The first allows users to specify the behavior of one or several actuators. In this case the designer has to use the same device as the end-user. In the second category the designer edits haptic cues that the user will perceive without referring to specific hardware. Various data formats and graphical tools are summarized in Table 2.2.

2.1.4.1 Device-oriented effects

The behavior of an actuator is typically controlled by specifying a curve representing the amplitude of the stimulation (vibration or the force in time). The Hapticons editor [EM03] was created to edit trajectory patterns called “haptic icons” on a 1DoF force feedback device (see Figure 2.3). Similar systems have been developed for 3DoF devices [GMS06, CSKR07]. In the same way, vibration patterns may be edited. This kind of tool is already used in the industry. The aforementioned Integrator [IMM] development platform provides a curve editor for designing vibrotactile patterns for various devices (mobile phones, gamepads, etc.).

Quite different graphical interfaces are used to edit the behavior of an array of motors. The user must specify the behavior of each motor in time. Representative examples have been developed by Rahman et al. [RAC10] and Kim et al. [KCRO10] (see Figure 2.4).

2.1.4.2 User-oriented effects

The second type of graphical tool focuses on describing what the user should feel instead of defining how actuators should behave. This implies that the haptic rendering is handled by dedicated software.

Ryu et al. have created the posVib Editor to edit vibrotactile patterns [RC08]. The intensity of the vibration felt by the user is represented by a curve.

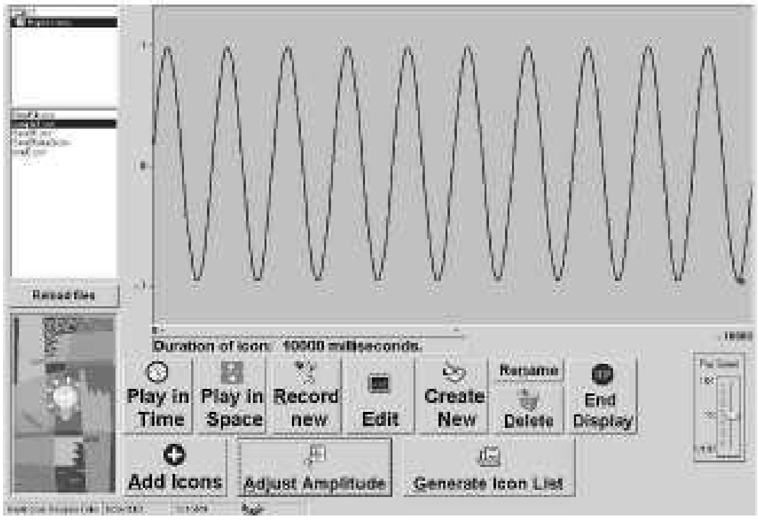


Figure 2.3: Device-oriented graphical authoring tools - Hapticons Editor [EM03].

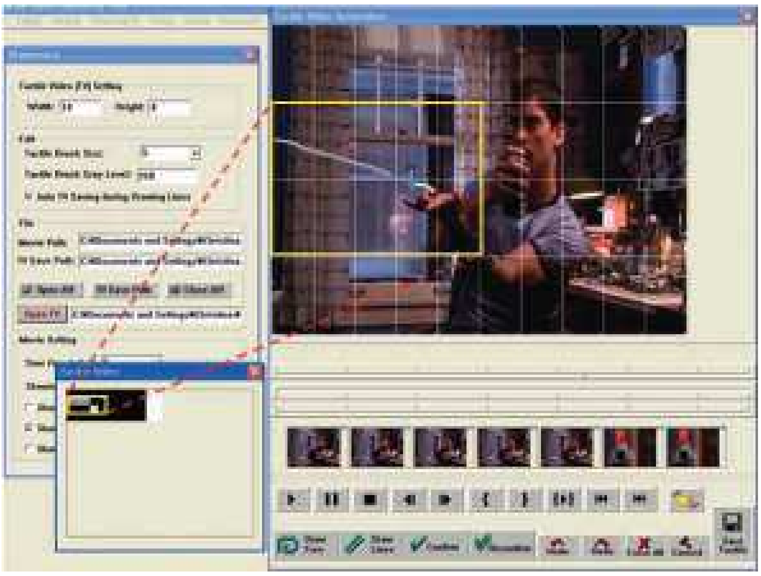


Figure 2.4: Device-oriented graphical authoring tools - Tactile array editor [KCRO10]).

Three editors, based on the MPEG-V format, have been developed to create and tune sensory effects all along a movie: Rose Studio [CLY11], SEVino [WRTH12] and SMURF [Kim13]. One or several effects can be added on a timeline which determines when they start and when they finish (see Figure 2.5). The haptic effects supported by these editors are vibrations, temperature, wind and water-spray.

A different approach consists in describing material properties of objects within a scene. It implicitly determines what users feel when they touch objects. This type of tool resembles a 3D editor in which the author directly visualizes the 3D object being manipulated, but haptic (friction, stiffness) rather than visual properties are edited. We refer the readers to the presentation of the K-Haptic Modeler [SLK+07] as well as the HAMLAT tool [EAAE08] which is a graphical editor for HAML (see Section 2.2.1.1 and Figure 2.6).

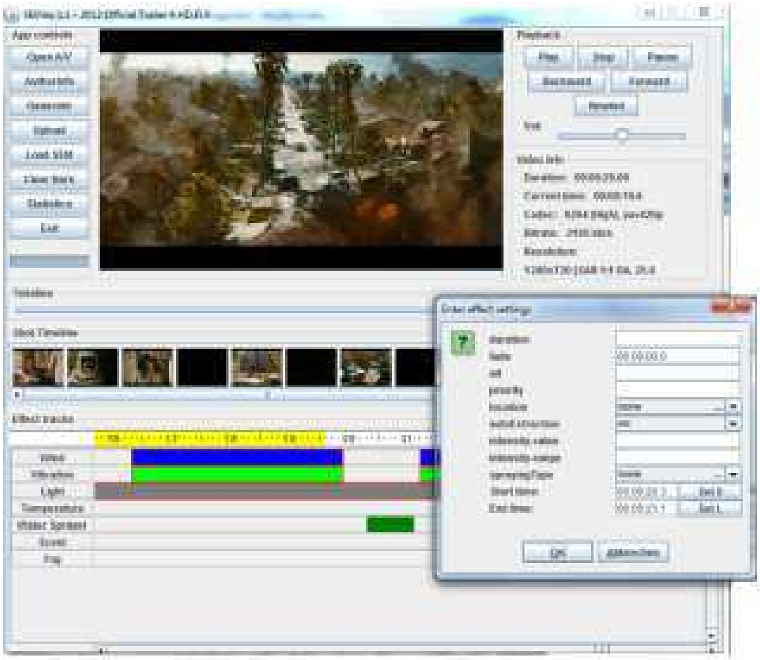


Figure 2.5: User-oriented graphical authoring tools - SEVino [WRTH12]. Haptic effects can be defined and synchronized to a video.

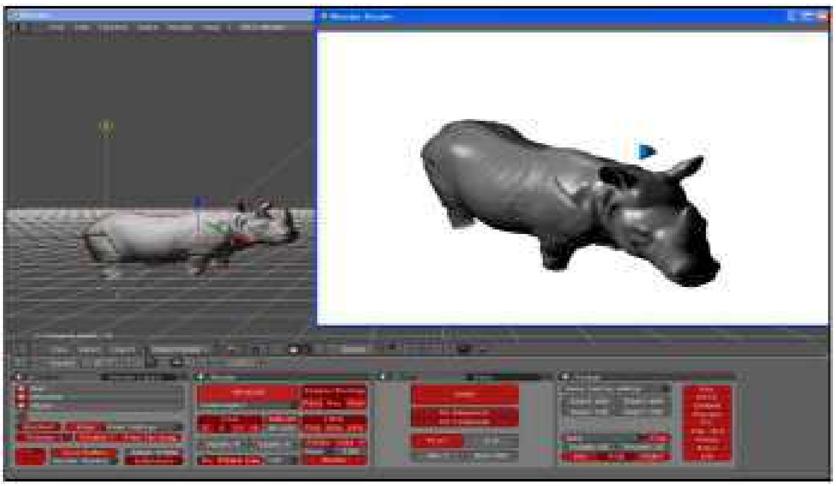


Figure 2.6: User-oriented graphical authoring tools - HAMLAT [EAAE08]. Haptic properties can be set for a virtual object.

Type of Effect	**Format**	**Data Content**	**GUI**	**Container**	**Ref.**
User-oriented	MPEG-V (XML)	Description and organization of sensory effects in a multimedia content	Yes (Rose Studio, SEVino, SMURF)	MPEG-2 TS	[CLY11, WRTH12, YCLL10, Kim13]
	MPEG-7 (XML)	Description of a 3D scene, haptic device and haptic rendering	Yes (HAMLAT)	n/a	[EfEOS06, EAAE08]
	XML	Haptic properties of a 3D scene: friction, stiffness, etc. of objects	Yes (K-Haptic Modeler)	n/a	[SLK$^+$07]
		Vibration patterns	Yes (PosVib Editor)	n/a	[RC08]
	VRML	Description of 3D objects and associated haptic rendering methods	No	n/a	[SW08]
	MPEG-4 BIFS (VRML)	Information about depth, stiffness, friction of a scene	No	MPEG-4	[CES09]
Device-oriented	n/a	Trajectory patterns	Yes (Hapticon Editor)	n/a	[EM03]
	XML	Description of haptic device properties and description of how they are activated	Yes (TouchCon)	n/a	[KSH09]
		Vibration patterns of a tactile array	Yes	n/a	[RAC10]
	MPEG-4 BIFS (VRML)	Vibration patterns of a tactile array	Yes	MPEG-4	[KCRO10, KPK$^+$11]

Table 2.2: Overview of existing formats to edit and store haptic effects. Two types of haptic effect can be described: effects focused on what the user will perceive (user-oriented), and effects focused on how the actuators will behave (device-oriented). Most of the time a graphical user interface is designed to easily edit data. Some formats are to be embedded with a container enabling both audiovisual and haptic contents to be distributed via streaming platforms.

2.2 Distribution of haptic effects

The second stage consists in formalizing haptic effects into data to be synchronized, stored and transmitted with the audiovisual media. Even though the range and nature of haptic effects is not yet well defined, there have been several attempts at providing formalizations. These formats are summarized in Table 2.2 which displays, when available, the associated authoring tools (see Section 2.1.4), and solutions to transmit haptic effects over the network (see Container column of Table 2.2).

2.2.1 Data formats for haptic effects

Though there are several contributions which use dedicated formats to encode haptic feedback for audiovisual content, most approaches rely on generic formats. We consider two ways to formalize haptic effects: "device-oriented" that defines the actuators' precise behavior, and "user-oriented" that describes effects from the user's point of view. The formats presented in this section are however suitable for both usages. Choosing between them only influences the way in which the rendering stage has to be handled: device-oriented data are used to control haptic devices directly, but user-oriented data must be interpreted. Since there is no obvious way to classify the encoding of haptic effects, we will use a per-format classification. We will detail contributions based on XML, a versatile description language, and VRML, a language dedicated to descriptions of 3D worlds. These formats are summarized in Table 2.2.

The issue of formalizing haptic effects has been solved by companies such as D-Box [DBO] or Immersion [IMM] who have developed commercial solutions for rendering haptic effects along with audiovisual content. D-Box has created a proprietary language to add haptic effects to a movie, called D-Box Motion Code™. However, details of these formats are not currently available and the effects cannot be edited by the end-user.

2.2.1.1 XML-based

The first method of formalizing haptic feedback relies on XML language. The Haptic Application Meta-Language (HAML [EfEOS06]) is a generic format for describing haptic feedback which contains information about the haptic device, haptic rendering and visual rendering (see Listing 2.1). The purpose of this format is to be able to use any haptic interface with any virtual world, the system adapting the haptic feedback to the capabilities of the haptic interface used. This language is dedicated to virtual reality applications but it could be used to describe scenes in audiovisual content: objects and their location, geometry, haptic properties (stiffness, damping, friction), etc. This format respects the MPEG-7 standard which yields standardized tools to structure and organize descriptions of multimedia content [CSP02].

```
<HAML>
...
    <SceneDS>
        <Object>
```

```
            <Type>Mesh</Type>
            <Name>Cube</Cube>
            ...
            <Tactile>
                <Stiffness>0.8</Stiffness>
                <Damping>0.9</Damping>
                <SFriction>0.5</SFriction>
                <DFriction>0.3</DFriction>
            </Tactile>
        </Object>
    </SceneDS>
</HAML>
```

Listing 2.1: Example of an xml-based file (HAML [EOEC11]). Here, the haptic properties (stiffness, friction and damping) of a 3D cube are defined.

Closely related to video viewing, the MPEG-V format also relies on XML [MPE11]. This language is designed to add sensory effects to any multimedia content: movies, video games, web content, etc. Users can create groups of effects and synchronize them with other media (see Section 2.1.1 for the list of effects). For each effect the designer can specify at least its intensity and duration. However devices and techniques to render effects are not specified. If converting an intensity into vibrations is simple, the rendering of a forward movement over 2 meters with an acceleration of $30 cm.s^{-2}$ is less straightforward (see Listing 2.2). At the time of writing this thesis, this format is close to being standardized by the MPEG working group. First implementations may be found in the literature [Yoo13, WRTH13].

```
<sedl:SEM>
    <sedl:Effect xsi:type="sev:RigidBodyMotionType" activate="true" si:pts="1593000">
        <sev:MoveToward distance="200" acceleration="30"/>
    </sedl:Effect>
    <sedl:GroupOfEffects si:pts="1647000">
        <sedl:Effect xsi:type="sev:VibrationType" activate="true" intensity-range="0 100" intensity-value="10"/>
        <sedl:Effect xsi:type="sev:WindType" activate="true" intensity-range="0 100" intensity-value="5"/>
    </sedl:GroupOfEffects>
</sedl:SEM>
```

Listing 2.2: Example of an xml-based file (MPEG-V [MPE11]). Here a "Move Toward" effect is defined followed by a group of effects combining "Wind" effect and a "Vibration" effect.

In an approach dedicated to instant messaging applications, Kim et al. [KSH09] developed an XML-based format to exchange haptic feedback called "TouchCons". This allows users to send haptic messages such as vibration patterns or thermal effects. Two main files are used in this system. First, the Library XML describes a list of haptic messages and how they should be rendered (device used, intensity, duration). Second, the Device XML describes the available devices and associated capabilities. To send a message, the user chooses one from the Library XML file. When he receives a message, it is rendered according to the capabilities of the devices listed in the user's Device XML file. This framework could be used, instead of TouchCons, to describe haptic effects and then to send them to the end-user. The effects would be then rendered according to the user's devices configuration.

Finally XML representation can be used to determine the behavior of actuators directly. For example, Rahman et al. [RAC10] described vibration patterns of a vibrotactile array: the vibration intensity of each motor is specified in an XML file. This approach is simple but the effects described can be rendered only by a specific device.

2.2.1.2 VRML-based

A third method used to describe haptic content uses VRML/X3D. This language serves to represent 3D worlds and contains information needed by visual rendering systems. Sourin and Wei [SW08] proposed an extension of this language by adding haptic rendering techniques. One purpose of this language is to transmit virtual objects and their associated haptic rendering algorithms over the internet. In a similar way to HAML, this solution allows an audiovisual scene and the associated rendering techniques to be described.

The two techniques presented hereafter are based on the MPEG-4 BIFS format, also known as MPEG-4 Part 11[ISO05]. BIFS, which stands for Binary Format for Scenes, is a scene description protocol based on VRML. Cha et al. extended this format to add haptic properties to a video [CES09]. The authors built a "touchable" movie, i.e. a movie in which spectators can feel the depth of the images using a force-feedback device. For each frame of the video the authors associated texture properties (stiffness, static friction and dynamic friction; see Listing 2.3).

```
Shape{
  appearance Appearance {
    texture ImageTexture {
      url "color_image.jpg"
      }
    hapticSurface HapticTextureSurface {
      stiffnessRange 0.1 10
      staticFrictionRange 0.2 0.9
      dynamicFrictionRange 0.3 0.9
      maxHeight 1.0
      hapticTexture ImageTexture{
        url "haptic_image.jpg"
        }
      }
    }
  geometry Depth {
    focalLength 6.983
    pixelWidth 0.00123
    nearPlace 10
    farPlane 200
```

```
    texture ImageTexture {
      url "depth_image.png"
      }
    }
}
```

Listing 2.3: A VRML-based file (Extended MPEG-4 BIFS [CES09]). This file describes haptic properties of a visual scene (color_image.jpg). The depth map and associated friction are specified.

The modified BIFS format can also be used to store vibrotactile patterns used to drive an array of vibration motors. Kim et al.'s encoded a pattern in a grey-scale image where each pixel represents an actuator and the intensity of the pixel corresponds to actuator activation intensity: from black (0) for idle to white (255) for maximal vibration [KCRO10]. In a similar way, vibrotactile patterns can be associated with video frames (see Listing 2.3: instead of "haptic_image.jpg" a "tactile_pattern.jpg" would be associated with the visual scene). Thus the MPEG-4 BIFS format extended by Cha et al. can both describe a 3D scene and/or contain data to drive vibrotactile arrays. These two possibilities have been implemented by Kim et al. for adding haptic textures effects or vibration effects to educational videos [KPK+11].

2.2.2 HAV containers

A container is a meta-file format that can hold several files in a single file or stream which makes distribution easier. In the HAV context, a container regroups haptic, visual and audio content. This stage is depicted in Figure 2.1. All components are compressed and synchronized into a single container for network transmission [WHZ+01]. These containers are mainly used in multimedia applications to store both audio and visual content into a single file which is then transmitted, downloaded or streamed.

Several containers embedding audio and video exist (ogv, avi, mp4, etc.), but those combining haptic content are less common. A simple solution would consist of directly embedding the file containing the haptic data into a container that allows the attachment, such as the mkv container. O'Modhrain and Oakley used the Flash standard to distribute videos enhanced with haptic effects [OO04]. They integrated haptic feedback in their home-made animation and the media was played by a web browser embedding the Immersion Web plug-in. This alternative is suitable for distribution purposes, although limited to the rendering capability of the plug-in and to a specific type of audiovisual content (animation).

To take advantage of streaming platforms, one solution is to develop formats for haptic effects compatible with video containers that permit playback as they are downloaded. Some formats were designed to support this streaming feature (see Section 2.2.1). Modified MPEG-4 BIFS [CES09] can be embedded into a classical MPEG-4 container. In a similar way MPEG-V is compatible with the MPEG-2 TS container [Yoo13]. Actually any MPEG-based format is compatible with any MPEG-based container. Moreover MPEG-V provides a binary representation of the sensory effects to enable a fast and efficient transmission. This streaming challenge has been identified as **haptic broadcasting** by Cha et al. [CHK+09]. This is a specific challenge different from the classical transmission of data for teleoperation [HS06]. The purpose is not to control a device remotely but to send multimedia containing audio, video and haptic content. The two formats presented are at an early stage of development but demonstrate the possibility of haptic broadcasting.

2.3 Rendering of haptic effects

Once the haptic content has been transmitted to the end-user, the haptic device needs to decode and render the content to provide the appropriate effect (in the same way that video is displayed on the screen or audio is rendered on the speakers). Here we review a list of haptic interfaces proposed for "enhanced" video viewing.

We classified these devices into four categories: wearable devices, handheld devices, desktop devices and haptic seats. The results are presented in Table 2.3.

2.3.1 Wearable devices

Wearable devices are designed to be worn by as the user experiences audiovisual content. Typically they are composed of several vibrotactile actuators embedded into clothes, as detailed in Rahman et al. [RAC10] (see Figure 2.7a). This topic has been intensively studied for virtual reality purposes [LYNH06] and many devices have been designed.

(a) vibrotactile jacket and armband [RAC10] (b) vibrotactile gloves [KCRO10]

Figure 2.7: Wearable haptic devices.

In the HAV context, exploring the idea of enhancing live sports experience, Lee et al. proposed a device with vibrotactile sensations through an assembly of 7x10 vibrotactors attached to the user's forearm [LLC+05]. This prototype was used to render movements of the ball on the field during a soccer game. The tactile array was mapped to the field and vibrations were triggered at ball locations. According to the authors this device allows the user to better understand ambiguous game situations.

Kim et al. designed a tactile glove for immersive multimedia [KCRO10, KPK+11]. It contains 20 tactile actuators per glove (4 per finger). The gloves are wireless-controlled and produce vibrotactile patterns as the user watches a movie (see Figure 2.7b). These patterns were first created, then synchronized with the video.

A vibrotactile belt has been designed by Ooshima et al. to provide a feeling of being slashed [OHA+08]. They have used small speakers to generate vibrations that propagate inside the user's abdomen.

A tactile jacket has also been developed by Lemmens et al. [LCB+09]. They explored the influence of tactile devices on spectators' emotional responses, and designed a tactile jacket with 16 segments of 4 vibration motors covering the torso and the arms. Motors are activated following patterns related to specific emotions. For example, the feeling of love is enhanced by activating motors overlying the abdomen in a circular manner.

Palan et al. [PWN+] presented a vest with embedded vibration motors, solenoids and Peltier elements. The vest was designed to display three haptic effects as realistically as possible: gunshots, slashing and blood flow, with the motivation of improving video games experience. Similarly, a commercially available jacket manufactured by TNGames produces effects such as explosions, gunshots or accelerations using 8 air cells [TNG].

While the embedded devices do not yield a significant change in weight or wearability of clothes, being composed of simple vibrotactile actuators, the range of possible haptic effects is rather limited.

2.3.2 Handheld devices

Users can experience haptic feedback through portable devices held in the hand. Vibrotactile technology appears well-suited for portable devices. For years, the gaming industry has used vibrating joypads to enhance immersion in video games. Mobile devices (phones and tablets) are now equipped with vibration motors which may be used to enhance multimedia contents [IMM]. Using this technology, Réhman et al. relied on a mobile phone equipped with a vibration motor to display haptic cues related to a soccer game [uR08]. Vibrotactile capabilities of tablets can be extended by a tactile surface, allowing then to touch images [GALSC12, KIP13]. Such systems are not used in a HAV context though. Alexander et al. developed a prototype of a mobile TV providing tactile feedback using ultrasound [AMS11]. The device is a screen with a 10x10 array of ultrasonic transmitters set on the reverse side. The user holds the device to observe the audiovisual content and experiences haptic feedback through the fingers (see Figure 2.8).

The remote control developed by O'Modhrain and Oakley is a different sort of handheld device that provides force-feedback [OO04]. A gaming joystick was rehoused in a device resembling a remote control (see Figure 2.9). Similarly Yamaguchi et al. used a computer mouse with a 2DoF force-feedback joystick [YAMS06].

As with clothes-based devices, handheld devices cannot embed heavy actuators and so only a restricted range of haptic effects can be rendered. However, the use of a common device in the user living space (remote control, mobile phone) seems well on the way to popular acceptance.

Figure 2.8: Handhled haptic device. A mobile haptic TV embedding an array of ultrasonic transmitters delivering tactile cues on user's fingers [AMS11].

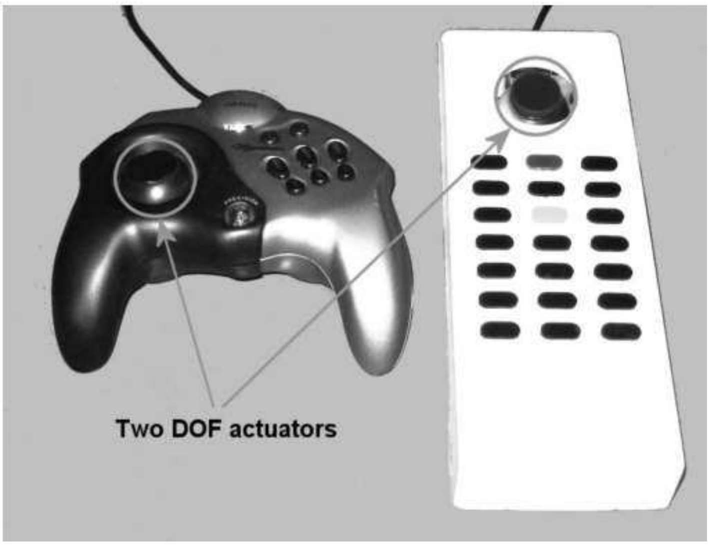


Figure 2.9: Handhled haptic device. A remote control including a 2DoF force-feedback joystick [OO04].

2.3.3 Desktop devices

In virtual reality settings, force-feedback devices are mainly used to interact with virtual objects. The user can feel and often modify the displayed content. With video viewing the user cannot modify the content. The user receives haptic cues, sometimes while actively exploring the content, but the audiovisual content does not change. For example in the solution devised by Gaw et al. [GMS06], the user holds a force-feedback device and is guided along a prerecorded path while viewing a movie. The same technique was used by Kim et al. to enhance educational videos with a Phantom device [KPK+11, PHA].

These devices have also been adapted to the task of "touching" images in a video [CES09] (see Figure 2.10). In this study the user could actively explore the video content and received haptic feedback through a Novint Falcon device [NOV].

Figure 2.10: Desktop haptic device. User touching a video thanks to a force-feedback device [CES09].

Other desktop devices have been designed to convey haptic feedback to the user without direct contact. An example is a fan which generates air streams, simulating the haptic effect of wind. Associated with a thermal device, a fan may be used to create temperature variations [Dio97]. Fans providing wind effects are commercially available. The Philips amBX system generates not only wind effects but also lighting effects and enables keyboard vibration [AMB]. This kind of device is simple to use, which results in more ecological interaction. Waltl et al. have relied on this device to enhance the multimedia experience [WRTH13].

Contact with virtual objects is possible without directly handling a device. Hoshi et al. [HTIS10] used ultrasound to exert pressure remotely on a user's skin. Their prototype was composed of an array of 324 airborne ultrasound transducers, able to exert a force of 16 mN at a 20 mm focal point diameter over a 180×180 mm surface. This invisible surface is created at 200 mm above the device. Combined with a 3D display system, the author succeeded in creating touchable floating images. A similar system has been developed by Sodhi et al. [SPGI13], based on the projection of air vortexes (see Figure 2.11). This technique allows a larger workspace than the previous device. Tactile feedback can be provided with a 75 degrees field of view, and within an 8.5 cm resolution at 1 meter.

Figure 2.11: Desktop haptic device. The AIREAL, a contact free tactile interface [SPGI13].

2.3.4 Haptic seats

Our fourth device category is the haptic seat. The user sits on a modified chair and passively senses haptic effects.

Vibrotactile actuators have once again been used in a number of ways. The tactile blanket [DWAD10], a variant for the theme Lemmens' Jacket [LCB+09], is equipped with 176 actuators and displays vibration patterns designed to enhance the user's emotion. More recently Israr and Poupyrev embedded an array of 12 vibrotactile actuators in the back of a chair, with an original controller [IP11]. The user experienced the tactile illusion of a continuous stimulus though the actuators were at discrete locations (see Figure 2.12).

Figure 2.12: Haptic seat. Seat embedding an array of vibrotactile actuators [IP11].

Several commercial products in this category are already available. One example is the "couch shaker" from The Guitammer Company [BUT]. This device uses actuators to shake the couch or sofa, operating like a subwoofer which propagates low-frequency vibrations to the couch instead of playing sounds. Some seating devices attempt to provide more complex effects such as motion. Typically such seats are fixed on actuators or motion platforms. For example, the D-Box seat features 3DoF: pitch, roll and heave [DBO] (see Figure 2.13).

Haptic seats are commonly encountered in theme parks or amusement arcades where they are typically used as motion simulators. Some of them even embed several devices to provide a wide range of effects (water spray, air blast, leg ticklers, etc. See the CJ 4DXPlex company [CJ4]). These devices are not, however, adapted to the end-user living space and their cost is prohibitive for the mass market. In contrast, the D-Box seat is a consumer product designed for living room use though it remains expensive. Devices based on vibrotactile arrays are also available but the range of tactile effects which can be rendered is quite limited.

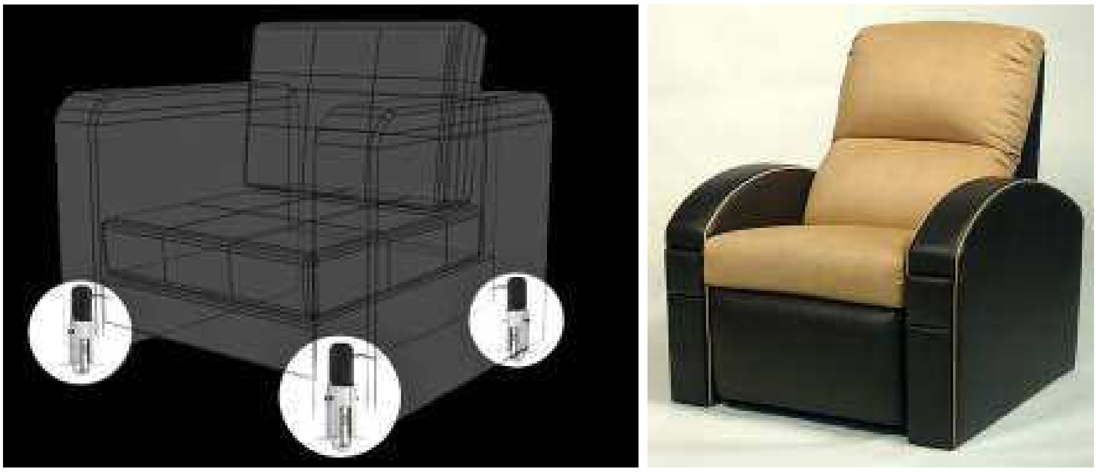


Figure 2.13: Haptic seat. The D-Box seat [DBO].

2.4 Quality of experience

Haptic effects aim at enhancing the audiovisual experience. This means that the quality of experience (QoE) of a video viewing session with haptic feedback would be higher than when haptic feedback is not present. But how should this hypothesis be assessed? Jain discusses the necessity of capturing the QoE for system evaluation [Jai04]. He underlines the difficulty of identifying and measuring the factors that characterize this metric due to its subjective nature.

Nevertheless Hamam et al. [HESG08] have proposed an initial model for the evaluation of QoE in multimedia haptics which identifies four factors: rendering quality, and the user-centered measures of physiology, psychology and perception. The rendering quality is dependent on the quality of the visual, audio and haptic feedback. Perception measures describe the way the user perceives the system depending on the user's expe-

Type of interface	**Device**	**Actuator**	**Haptic Effect**	**Ref.**
Wearable	Vibrotactile armband	7x10 vibration motors	Vibrations (related to position of a ball during a soccer game)	[LLC+05]
	Vibrotactile glove	20 vibration motors (4 per finger)	Vibrations	[KCRO10]
	Vibrotactile armband or jacket	Array of vibration motors (variable size)	Vibrations	[RAC10]
	Vibrotactile jacket	16x4 vibration motors	Vibrations (related to user's emotions)	[LCB+09]
	Vibrotactile vest	Vibration motors + solenoids + peltier elements	Pressure (gunshot), temperature (blood flow), vibrations (slashing)	[PWN+]
	Vibrotactile vest	8 air cells	Vibrations and pressure (gunshots, acceleration, explosion)	[TNG]
Handheld	Mobile phone	Vibration motor	Vibrations (related to status of soccer game)	[uR08]
	Mobile phone	Vibration motor	Vibrations	[IMM]
	Remote control	2DOF joystick	Force	[OO04]
	Computer mouse	2DOF joystick	Force	[YAMS06]
	Portable TV	10x10 array of ultrasound transducers	Pressure	[AMS11]
Desktop	Force-feedback device	3DOF motorized arm	Movement	[GMS06]
	Phantom	6DOF motorized arm	Movement	[KPK+11]
	Novint Falcon	3DOF motorized arm	Force (texture of an image)	[CES09]
	n/a	Array of 324 ultrasound transducers	Pressure	[HTIS10]
	n/a	Air vertexes projector	Pressure	[SPGI13]
	Philips AmBX	Vibration motor + 2 fans (+ 2 LED spotlights)	Vibration (+ wind & light)	[WRTH13]
Haptic Seat	Vibrotactile blanket	176 vibration motors	Vibrations (related to user's emotions)	[DWAD10]
	Vibrotactile chair	3x4 vibration motors	Vibrations	[IP11]
	Couch	Vibration motor	Vibrations (of the whole seat)	[BUT]
	Moving chair	4 compressors under chair legs	3DOF body motion (pitch, roll, heave)	[DBO]

Table 2.3: Overview of existing haptic devices used for enhancing audiovisual content.

rience, fatigue and other factors which may alter the user's perception. Physiological measures identify how the system modifies the user's biological state, and psychological measures highlight changes in mental state. The authors detail an exhaustive list of parameters related to each factor (e.g. respiration rate, body temperature or blood pressure for physiological measures). While this provides a taxonomy of the different factors influencing the quality of experience, techniques to evaluate them were not presented.

In this section we detail classical techniques to measure the QoE of HAV systems. The typical approach found in the literature is a subjective measure based on questionnaires. Other approaches capture biosignals which provide an objective measurement of the user's physiological state from which emotional state is inferred.

2.4.1 Subjective measures: questionnaires

Most contributions in HAV rely on simple questionnaires to evaluate the impact of haptic feedback on the quality of experience. Participants are usually asked to respond to questions on a Likert-scale. For example, Kim et al. [KCRO10] studied the benefits of vibrotactile feedback for enhancing movies by using four general questions (Is this more interesting than movies? Is the tactile content easy to understand? Is the tactile content related to the scene? and Does the tactile content support immersion?). Ur Rhéman et al. covered the same aspects using a more detailed questionnaire [uR08].

A more elaborate approach characterizes the quality of experience using multiple factors. Hamam et al. [HGE10] evaluated the five factors (extracted from their model described above) of realism, usefulness, intuitivism, fatigue and QoE.

Waltl et al. have evaluated the QoE by presenting a video two times to participants, first without sensory effects and then the video augmented with sensory effects. Participants were asked to quantify on a Likert-scale the enhancement brought by the effects [WT10].

The variation of approaches highlights the need for a standardized questionnaire to better evaluate and compare different systems. Identifying the factors to be measured is not an easy task, but several have already been evaluated in a systematic way: comfort, interest, acceptance and satisfaction. They can serve as a basis on which to build a subjective measure of the QoE.

2.4.2 Objective measures: physiological data

Another approach to the evaluation of the quality of experience consists of measuring changes in the user's physiological state. The QoE cannot be directly determined from this measure, but it can be used to infer the user's emotional state, which contributes to the QoE. To the best of our knowledge, no work has been done using these techniques in the context of HAV systems. Nonetheless, inspiring results can be found in the context of virtual reality applications and video viewing.

In the context of virtual reality, Meehan et al. gathered heart rate, skin conductance and skin temperature data from subjects in a stressful virtual environment [MIWB02].

These measures helped to determine the user's feeling of "presence" and were compared to subjective users' self-reports (see [SVKV01] for a survey on "presence"). These authors suggest that heart rate has the strongest correlation with a sensation of presence. Skin conductance correlated less strongly and skin temperature not at all. Haptic feedback significantly improved presence.

Mandryk et al. observed biosignals in video game players to determine their user experience [MIC06]. Skin conductance, heart rate, facial muscle activity and respiration rate were captured. The authors concluded that, for most participants, playing against a friend is more enjoyable than playing against the computer. The physiological measures were significantly consistent with the self-reported measures.

In a video viewing context, Fleureau et al. studied the potential of physiological signals for detecting emotional events [FGHT12]. Participants simply watched several videos while their heart rate, skin conductance and facial muscle activity were recorded. A detector based on machine learning techniques was designed. Given the user's biosignals, the system was robustly able to determine whether users were experiencing an emotional event and if this event was positive or negative.

The physiological chosen signals in these studies were mostly similar: heart rate, galvanic skin response, and facial muscle activity. All yielded significant results despite the various settings of virtual reality, video games and video viewing. The implications for the evaluation of HAV experiences are clear. Furthermore, closed-loop systems, in which physiological signals are used to control the nature and intensity of haptic events offer interesting possibilities for adapting the haptic effects to the individual user.

2.5 Discussion

We have presented an overview of how haptic effects can enhance audiovisual content. Studies relevant to each stage of haptic production, distribution and rendering have been presented. Some of these studies present solutions that address all stages and may be seen as implementations of the generic workflow displayed in Figure 2.1. These general approaches are summarized in Table 2.4.

While the existing solutions clearly demonstrate how haptic effects can be used with audiovisual content using tactile or kinesthetic feedback, the studies reported do not explore combinations of effects (e.g. kinesthetic and tactile). This is mostly because the devices studied have generally had only one type of actuator. As a consequence, the range of effects that can be generated is narrow and the conjunction of effects is rarely explored, despite the significant potential benefits. Furthermore, there appears to be a gap between the use of portable haptic interfaces (wearable or handheld), conveying weak effects, and complex devices (motion simulators) which are not adapted to the user living space. There is thus a clear opportunity to design new haptic devices dedicated to audiovisual enhancement. This implies in turn a better understanding of the requirements for HAV systems, which seem to differ significantly from those in virtual reality systems.

Further research on user perception should be conducted to determine relevant haptic stimuli for effective audiovisual entertainment. The link between haptic stimuli and user experience is not thus far well established. Haptic effects are mainly used in a similar way to the use of haptic feedback in virtual reality: to immerse the user physically in the audiovisual scene. The use of haptic effects to enhance non-diegetic aspects of a video such as the ambiance or emotions has been little studied. This appears as a key challenge and opportunity in this nascent field.

The distribution stage also requires research effort. Each solution currently uses a different technique to formalize haptic effects in the absence of a common definition for haptic effects. Only half of the studies have proposed methods for the transmission of the media to a remote display device. But several techniques allowing haptic broadcasting are emerging. Multimedia containers embedding audiovisual and haptic effects are currently being developed and standardized (MPEG-V, MPEG-4 BIFS). The MPEG-V format is a promising standard for distribution currently under development by the MPEG group. The draft standard presents a list of haptic effects along with an XML-based method to describe them. This format is also designed to be compatible with streaming technologies. However the new standard will have to follow the evolution of this emerging field of study. New haptic effects and devices will almost certainly be developed.

In most solutions haptic effects are synthesized: authors manually create and synchronize haptic effects to the audiovisual content. Each solution currently offers a different technique for editing haptic effects, though general editing tools may arrive with the advent of new standards. The automatic extraction of haptic cues from visual content has also been reported. Such cues are currently limited to specific audiovisual content: soccer games following pre-defined rules, and animations where the position and geometry of objects is already known. The automatic extraction of haptic effects for any audiovisual content remains a complex task, and more work will be necessary to adapt current algorithms to this new purpose. Extraction can be facilitated by metadata that describe the content of the media, but extracting haptic effects from videos is a new challenge for which new specific techniques will have to be designed.

One final aspect to be discussed in this review is the quantification of the benefits lent to audiovisual content by haptic effects. Some of the studies presented here have conducted user evaluations, mostly based on questionnaires. Most show that haptic effects enhance the user experience but the various studies are heterogeneous and hardly comparable. There is pressing need for shared validated tools to evaluate this quality of experience.

2.6 Chapter conclusion

In this chapter we have surveyed the possibilities provided by haptic feedback for enhancing audiovisual content. Several trends can be identified within this emerging field. The studies presented have been organized within a workflow and the key challenges that pertain to this new way of experiencing videos identified.

The first stage of the workflow, related to production of haptic effects, is the iden-

Audiovisual Content		Haptic Effect	Production	Distribution	Rendering	Ref.
Category	Details					
Sport	Soccer game (3D simulation)	Vibrations (ball position)	[Automatic extraction] The system traces the ball during soccer game	n/a	Vibrotactile array embedded into an arm band	[LLC+05]
	Soccer game (simulation)	Vibrations (ball position, goals, team leading)	[Automatic extraction] Video analysis of events (not implemented, events are received from the simulation)	n/a	Mobile phone equipped with vibration motor	[uR08]
Animation	Animation (home-made with Flash)	Force (related to an object of the animation)	[Automatic creation] Force-feedback is computed from the position and geometry of the object	Flash	Mouse with a joystick (2DOF force feedback)	[YAMS06]
	Cartoon (home-made with Flash)	Force (related to on-screen character)	[Synthesis] Force-feedback is defined during edition of the cartoon	Flash	Remote control with a joystick (2DOF force feedback)	[OO04]
	Cartoon / Movie	Movement (user's hand is guided)	[Capturing] Trajectories recorded from force feedback device	n/a	Force-feedback device	[GMS06]
Movie	Movie	Force (user touches the image)	[Synthesis / Capturing] Material properties for each frame (depth, stiffness, etc.) stored into MPEG-4 BIFS	MPEG-4	Novint Falcon (3 DOF force-feedback)	[CES09]
	Movie (from Youtube)	Vibrations	[Synthesis] Vibration patterns stored into XML file	XML file on a web server	Vibrotactile array embedded into arm band or jacket	[RAC10]
	Movie	Vibrations	[Synthesis] Vibration patterns stored into MPEG-4 BIFS	MPEG-4	Vibrotactile array embedded into gloves	[KCRO10]
	Movie	Vibrations and wind	[Synthesis] Sensory effects stored into MPEG-V file	MPEG-2 TS	Philips amBX system	[WRTH12]
	Educational video	Vibrations or force (user touches the image)	[Synthesis] Haptic effects (vibrations or haptic properties) stored into MPEG-4 BIFS	MPEG-4	Vibrotactile gloves or Phantom device (6DOF force-feedback)	[KPK+11]

Table 2.4: Summary of existing schemes for adding haptic effects to audiovisual content. Each system offers a solution for synchronizing and rendering haptic feedback within an audiovisual content. Some schemes also specify ways to distribute the media over the network.

tification and generation of haptic effects which must be delivered to the user during the display of the media. We detailed different formats to store and synchronize haptic effects to audiovisual media, from VRML-based representations to standardized XML formats. The key issue is the creation of haptic feedback. While a number of authoring tools are available, these effects may also be captured from physical sensors or generated from another part of the media (video, audio or metadata).

Once the media has been enriched with haptic effects, it must be sent to the user. Media streaming platforms to distant users is now a common method of distribution. This second stage is dependent on the way haptic data are stored. Though these issues are largely solved for audiovisual media, there are few standards for media with haptic effects. However some pioneering contributions have demonstrated the feasibility of this approach.

In the last stage the user perceives the media through a haptic device. These haptic interfaces are generally designed and dedicated to the purpose of displaying haptic cues during video viewing. They may be portable (wearable or handheld devices) or grounded (desktop devices or haptic seats).

Each stage impacts the user experience. The evaluation of the quality of experience is then a key challenge of the field, but few methods exist. The subjective experience may be evaluated though questionnaires and more objective measures can be collected from physiological data.

The results of our survey suggest that research effort is needed in the design of data formats and technology for distributing HAV content. Promising solutions are currently under development. The development of haptic media creation tools is also necessary. This may lead to a new type of professional activity in the cinema industry. Just as 3D movies now need "stereographers", so will new HAV content require "haptographers". Haptic devices adapted to video viewing settings are also needed to render this new content. Moreover the development of tools to evaluate the quality of experience and the acceptance of such systems is mandatory. Tackling the challenges of this young but promising field of study will yield new tools and methods for adding haptic content to multimedia, leading to a more compelling user experience in combination with audiovisual content.

Part I

Rendering Haptic Effects: Novel Device and Algorithms for Rendering Haptic Effects in Video Viewing Settings

Chapter 3

HapSeat: simulating motion sensation with multiple force-feedback devices embedded in a seat

Contents

Motion simulation is usually provided by a motion platform [LD03]. Typically the user's whole body is moved to generate various sensations such as accelerating, falling or passing over bumps. While these devices generate a realistic sensation of motion with 6 degrees of freedom (DoF), they are not designed for domestic settings and they are too expensive for the mass consumer market. Immersive experiences with motion effects are thus currently limited to amusement parks or "4D Cinemas".

In this chapter we introduce the *HapSeat*, a novel approach for producing motion sensations in a consumer settings using multiple force-feedback devices embedded in a seat. We first review the literature on human motion perception and motion simulators in section 3.1. Then the *HapSeat* is introduced in section 3.2. Three low-cost actuators which simulate 6DoF effects of motion are used, the motion effect is generated by adjuncts to the structure of the chair rather than moving the whole chair. The proof-of-concept is detailed in section 3.3. A prototype has been designed and constructed, which uses actuators held by an armchair-shaped structure. Two models to control the device have been implemented: a *Physical Model* which computes forces supposed to be felt during a movement, and a *Geometrical Model* which modifies the structure to match the position and posture that characterize a movement. A user study was conducted to assess this approach and to evaluate the quality of the user experience. Protocol and results are presented in section 3.4. Finally a conclusion is provided in section 3.5.

3.1 Related work on motion simulation

3.1.1 Human motion perception

The perception of motion is a complex sensation resulting from the integration of multiple perceptive inputs from different systems: visual, auditory, vestibular and kinesthetic [Ber00, HJZ+02]. The visual system contributes to this perception by providing an estimation of distances between the body and landmarks. A displacement of the body will modify these distances and add the perception of self-motion. Moving visual cues can often trigger a sensation of self-motion even though the viewer is stationary [RSP]. This illusion is called vection. The auditory system may also contribute to this perception by locating the body relative to "acoustic" landmarks [VLVK05].

The main contributor to the perception of motion is the vestibular system. Located in the inner ear, this organ is composed of three orthogonally-oriented semi-circular canals and two otolith organs. The canals allow rotational movements to be detected while otolith organs contribute to the perception of linear accelerations.

Additionally, it is interesting to note that haptic cues provided by the kinesthetic system also influence the sensation of motion. The kinesthetic system provides information about limb positions. When an elevator goes up, one feels the motion thanks to the proprioceptive receptors in joints and muscles of the legs. The tactile sense also provides information about motion: internal receptors detect movements of visceral organs and act as accelerometers. These visceral graviceptors are especially to be found in the region of the kidney. Similarly the somatosensory system indicates the direction of gravity through pressure patterns all over and inside the body [TBN+04].

3.1.2 Motion simulators

Motion simulators are well-known devices designed to make the user feel motion. They are intensively used in driving or flight simulators for training purposes. Most are based on a Stewart's platform [Das00], a 6DoF platform driven by 6 hydraulic cylinders. A motion simulator is basically a seat attached to this kind of platform. While the user navigates the virtual environment, the seat moves to generate a sensation of motion. These systems are often used in virtual reality rooms or 4D cinemas but few are designed for the mass market.

To the best of our knowledge, the D-Box company is one of the few actors in this market, having developed an armchair placed on four actuators that is suitable for an end-user's living-rooms [DBO]. This seat generates 3DoF motion effects (pitch, roll and heave) for movie viewing and consumer applications. Despite this attempt to succeed in the consumer environment, this chair remains expensive and limited to 3DoF motion effects.

The sensation of motion can also be induced in a less invasive way by force-feedback devices that simulate the kinesthetic system. Ouarti et al. applied a force to users' hands as they watched an optic flow stimulus [OLB09]. The system was expected to generate an illusion of motion with force-feedback: when the interface pulled the user's hand, the user experienced a sensation of forward motion. Similarly, Lécuyer et al. have showed that a torque feedback applied on a user's hand contribute to the feeling of self-motion [LVJ+04].

The use of haptic illusions to enhance the audiovisual experience has also been explored by Israr and Poupyrev, who designed a chair with several vibration devices embedded in the back [IP11]. Actuators in the chair were activated in such a way that the user felt a continuous stimulus. Though no effect of motion was claimed in this study, Riecke et al. have showed that vibrotactile feedback may generate a vection effect by improving the realism of the simulation [RSPCB05].

To sum up, there remains an important gap between haptic devices which do not, or only partially simulate, a sensation of motion, and complex simulators which are efficient in conveying motion but remain expensive and not well adapted to the consumer environment. We propose the *HapSeat* as a solution to fill this gap.

3.2 HapSeat: a novel approach for simulating 6DoF motion

We propose to enhance the experience of passive navigation in virtual or cinematic content using 6DoF motion effects generated by multiple force-feedback devices. Instead of moving the whole user's body as on motion platforms, only some parts of the body are stimulated. As described in section 3.1.1, the perception of motion results from the stimulation of various parts of the body (vestibular system, visceral organs, kinesthetic system). Our approach is built on the hypothesis that local haptic cues suffice to trigger a sensation of self-motion.

Using only one or two 3DoF force-feedback devices is not sufficient to invoke a 6DoF sensation of motion (translations and rotations). Up to 5DoF can be provided with two devices [STMA10]. Extending the approach to three 3DoF devices in order to apply three force-feedback stimulus to the user's body offers the possibility of simulating a global 6DoF effect of motion. A plane looping sensation could be simulated by pulling the head backward and lifting both arms simultaneously, while a car braking could be simulated by pushing both the head and hands forward. This concept can be extended by stimulating other regions of the body, using 5×3DoF devices for instance.

3.3 Proof-of-concept

The prototype developed as a proof of concept relies on three actuators. Two stimulate the user's hands, while a third stimulates the head. As the vestibular system is located in the head, stimulating this part of the body should heighten the illusion of simulated motion.

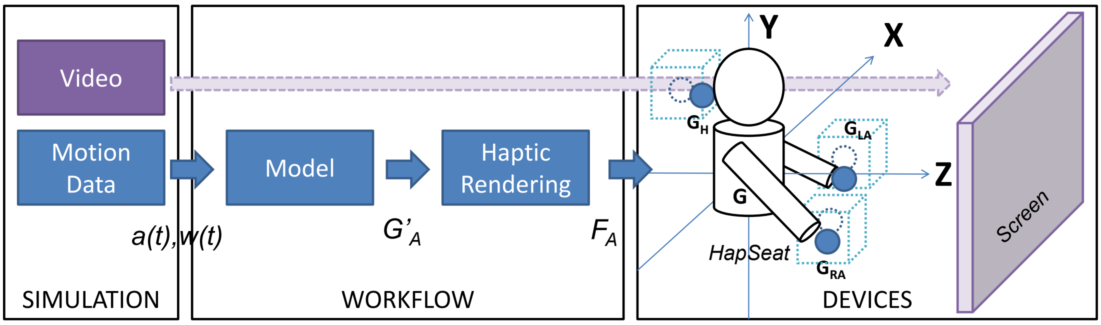


Figure 3.1: Simulating 6DoF motion with 3×3DoF force-feedback devices. While the three local devices are moving, the user is expected to feel a sensation of motion in relation to the visual content.

Figure 3.1 shows a schematic representation of the concept and offers an introduction to our notation. The motion description associated with a simulation is transmitted to a model at each instant t which computes the ideal position G'_A for each local actuator A. This position is then rendered by the haptic rendering algorithm as a force F_A. Each step of this workflow is detailed in this section.

3.3.1 Prototype of the HapSeat

An aluminum structure was designed to allow the positioning of the three actuators around an ordinary chair. The user passively rests his or her head and hands on each of the 3DoF actuators while watching a projection on a screen positioned in front of the chair (see figure 3.2). The head actuator is equipped with a block of foam for the user's comfort.

At rest (no rendered motion), the three actuators H, LA and RA maintain the head, right arm and left arm of the user at the central positions G_H, G_{LA} and G_{RA}

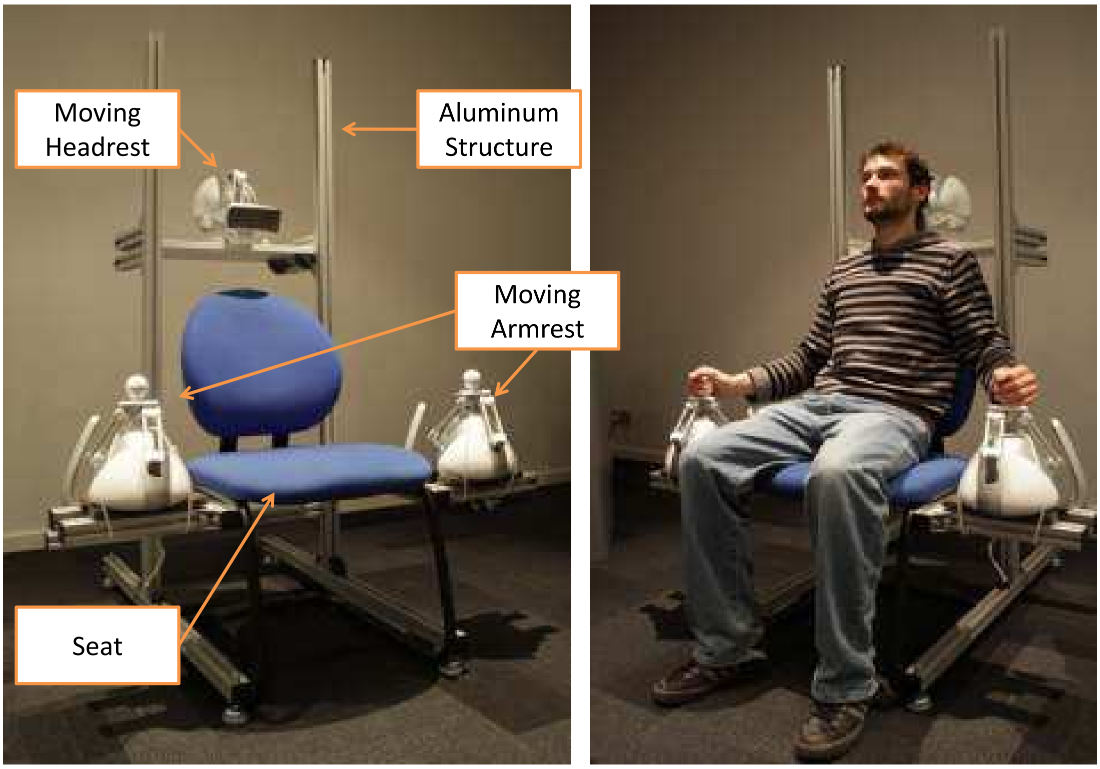


Figure 3.2: Prototype of the HapSeat. Left: seat structure with 3 force-feedback devices. Right: the system in use.

respectively. When the simulation starts, each actuator generates 3DoF forces on its respective body part within the limits of the cubic workspaces in Figure 3.1.

Our current prototype uses three Novint Falcons actuators [NOV]. These devices are robust, relatively cheap and the forces generated are appropriate for safe movement of the user's head and hands.

3.3.2 Motion data

We focus on the case of a first person point-of-view simulation, whose intention is to mimic for the user the sensation of motion that the principal actor would have felt at the time of the shooting. The audiovisual content is augmented with extra data describing the motion in terms of the linear acceleration $a(t)$ and the angular velocity $w(t)$. Let us define F_N as the navigation frame of the actor and F_B the frame associated with his body, centered on a point C (his chest for instance). The actor's motion is modeled as a rigid body motion described by two quantities $a(t) = [a(C \in F_B|F_N)]_{F_B} = \{a_x(t), a_y(t), a_z(t)\}^t$ (the gravity being removed [Sab06]) and $w(t) = [w_{F_B/F_N}(t)]_{F_B} = \{w_x(t), w_y(t), w_z(t)\}^t$ (where the $[x]_F$ notation designates the vector x expressed in the frame F).

This kind of content can be easily produced by a video camera equipped with an inertial measurement unit. The capture device was attached to an actor's torso to record both his movement and a video of his field of view (we describe such a system in Chapter 5). Therefore $a(t)$ and $w(t)$ describe the motion of the actor.

3.3.3 Models for motion simulation

Each actuator (H, LA, RA) moves to create the feeling of 6DoF global motion modeled by the quantities $a(t)$ and $w(t)$ as if the motion of the main actor was mapped onto them. Two models to control the device were devised. The first model is based on a *Physical Model*. The related acceleration applied to some parts of the body of the actor (here the head, left hand and right hand) are derived from the parameters of the global motion, $a(t)$ and $w(t)$ and then reproduced on the user by the corresponding actuators. The second model, referred to as *Geometrical Model*, aims at reproducing the position and attitude of the main actor on the basis of a more metaphorical paradigm.

3.3.3.1 Physical model

In this model the accelerations felt by the main actor at his head, P_H, and at his left and right shoulders, P_{LS} and P_{RS}, are computed through a rigid body approach, where the motions of the hands are considered equivalent to the movements of the shoulders. Knowing $a(t)$ and $w(t)$ at the origin of his body frame F_B, the accelerations of a new point P of the rigid body may be computed by the following mechanical relation (time derivation of the kinematic torsor):

$$[a(P \in F_B|F_N)]_{F_B} = a(t) + \frac{dw}{dt} \wedge \overrightarrow{GP} + w \wedge (w \wedge \overrightarrow{GP}) \tag{3.1}$$

The new position G'_A for an actuator A is formulated in terms of displacement from its initial and central position G_A by:

$$\overrightarrow{G_A G'_A} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{bmatrix} (a(t) + \frac{dw}{dt}(t) \wedge \overrightarrow{GP_A} + w(t) \wedge (w(t) \wedge \overrightarrow{GP_A})) \tag{3.2}$$

where G'_A is the new application point at time t, and s_x, s_y, s_z are scaling factors which map the actual motions of the three actuators in their workspaces. Those scaling factors are computed so as to use the workspace of the actuator in an optimal way. This involves compromises between the use of the largest possible space, so as to have a larger amplitude in the final rendering, while avoiding any saturation. These scaling factors are computed is a preprocessing step that consists of finding the maximal amplitude of the acceleration rendered by the actuator.

In this context the new application points G'_H, G'_{LA} and G'_{RA} are computed from the initial points G_H, G_{LA} and G_{RA}, and $s_x = s_y = s_z$.

3.3.3.2 Geometrical model

This second model aims to make the chair reproduce the position and posture of the moving actor during the simulation. Two kinds of motion will be rendered: linear accelerations and orientation changes. The linear acceleration rendering is simply performed by a simultaneous translation of each of the different local actuators along the 3D vector given by $a(t)$. The scene pose changes rendering is trickier. It makes the assumption that the rotation speed of the current scene, modeled by $w(t)$, may be rendered by rotating the position of the three actuators around the center modeled by a point G located near the user's sternum (see Figure 3.1) and with a 3D angle modeled by w. Then the faster the object is turning, the bigger the angle of rotation. Moreover, if the rotation stops (i.e. $w(t) = 0$), the actuators are at rest.

The new position G'_A of the actuator A for a rotation around G can be expressed:

$$\overrightarrow{GG'_A} = (R_x(w_x(t))R_y(w_y(t))R_z(w_z(t)))\overrightarrow{GG_A} \tag{3.3}$$

i.e.:

$$\overrightarrow{G_AG'_A} = \overrightarrow{GG'_A} - \overrightarrow{GG_A} \tag{3.4a}$$

$$= (R_x(w_x(t))R_y(w_y(t))R_z(w_z(t))) - I_3)\overrightarrow{GG_A} \tag{3.4b}$$

where R_x, R_y and R_z are the 3D rotation matrices around their respective X, Y and Z axes and I_3 is the identity matrix in dimension 3.

A complete 6DoF motion is a combination of linear accelerations and rotations. A function f is proposed to model the incorporation of both these types of information in our system. The proposed system has intentionally decoupled the linear motions from the rotational ones. This assumption is somewhat unrealistic from a mechanical point of view, but nevertheless makes sense in the context of passive navigation. If the motion to be rendered is a pure translation or a pure rotation, this decoupling is not a restriction. The difficulty arises when the motion to be rendered is a combination of translation and rotation. We make the assumption that a user would unconsciously expect to feel the dominant motion in the scene more strongly.

Then, if G'_A represents the new position of the actuator A at time t and G_A its initial position, we have:

$$\overrightarrow{G_AG'_A} = f(\begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{bmatrix} a(t), (R_x(m_xw_x(t))R_y(m_yw_y(t))R_z(m_zw_z(t)) - I_3)\overrightarrow{GG_A}) \tag{3.5}$$

with

$$f(B, C) = \frac{\|\vec{B}\|B + \|\vec{C}\|C}{\|\vec{B}\| + \|\vec{C}\|} \tag{3.6}$$

From this equation, the new application points G'_H, G'_{LA} and G'_{RA} are computed from the initial points G_H, G_{LA} and G_{RA}.

In addition, s_x, s_y, s_z on one hand and m_x, m_y, m_z on the other hand represent different scaling factors to map the actual motion represented by the couple $(a(t), w(t))$ in the workspace of the different actuators. As previously described, those scaling factors are computed so as to use the workspace of each actuator in an optimal way. More precisely, computing the scaling factors m_x, m_y and m_z is performed during a preprocessing step that consists in finding the maximal amplitude f with respects to the values of $a(t)$ and $w(t)$ over the whole time interval. An exhaustive numerical analysis is thus performed to find the joint optimal discretized values m_x, m_y and m_z. Several solutions may be admissible in the parametric space and the one that offers the best isotropic behavior is selected.

3.3.3.3 Output of the models

A comparison of the outputs of the models is described in this section to highlight their main differences. The outputs of simple translations and then rotations are combined together.

A linear forward acceleration on the Z axis can be described by $a(t) = \{a_x(t) = 0, a_y(t) = 0, a_z(t) = t\}^t$ and $w(t) = \{w_x(t) = 0, w_y(t) = 0, w_z(t) = 0\}^t$. Such a movement is rendered in the same way by both models:

$$\overrightarrow{G_A G'_A} = f(\begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{bmatrix} a(t)) \tag{3.7}$$

All actuators are moving simultaneously along the Z-axis as if the user is being pushed forward. The same behavior is observed for the other translations on Y and X axes. In these cases, the user is pushed upward or pulled toward the left side.

Secondly self-rotations are tested. For instance a left rotation around the Y-axis can be described by $a(t) = \{a_x(t) = 0, a_y(t) = 0, a_z(t) = 0\}^t$ and $w(t) = \{w_x(t) = 0, w_y(t) = \frac{t^2}{2}, w_z(t) = 0\}^t$ (the angular acceleration $w'(t)$ is linear). In this case (see Figure 3.3), the outputs of the models are different. With the *Physical Model* the user's hands are moving along the X-axis toward the center while the head is not moving. With the *Geometrical Model*, the right hand is going forward (Z-axis) and the left hand is going backward (Z-axis) while the head slightly moves to the right side (X-axis). The same behavior is observed for rotations on other axes: the *Physical Model* renders self-rotation by an attraction of each part of the body toward the center G and the *Geometrical Model* renders them with desynchronized movements.

A 6DoF movement that combines translations and rotations is thus managed differently by each model depending on the amount of rotations.

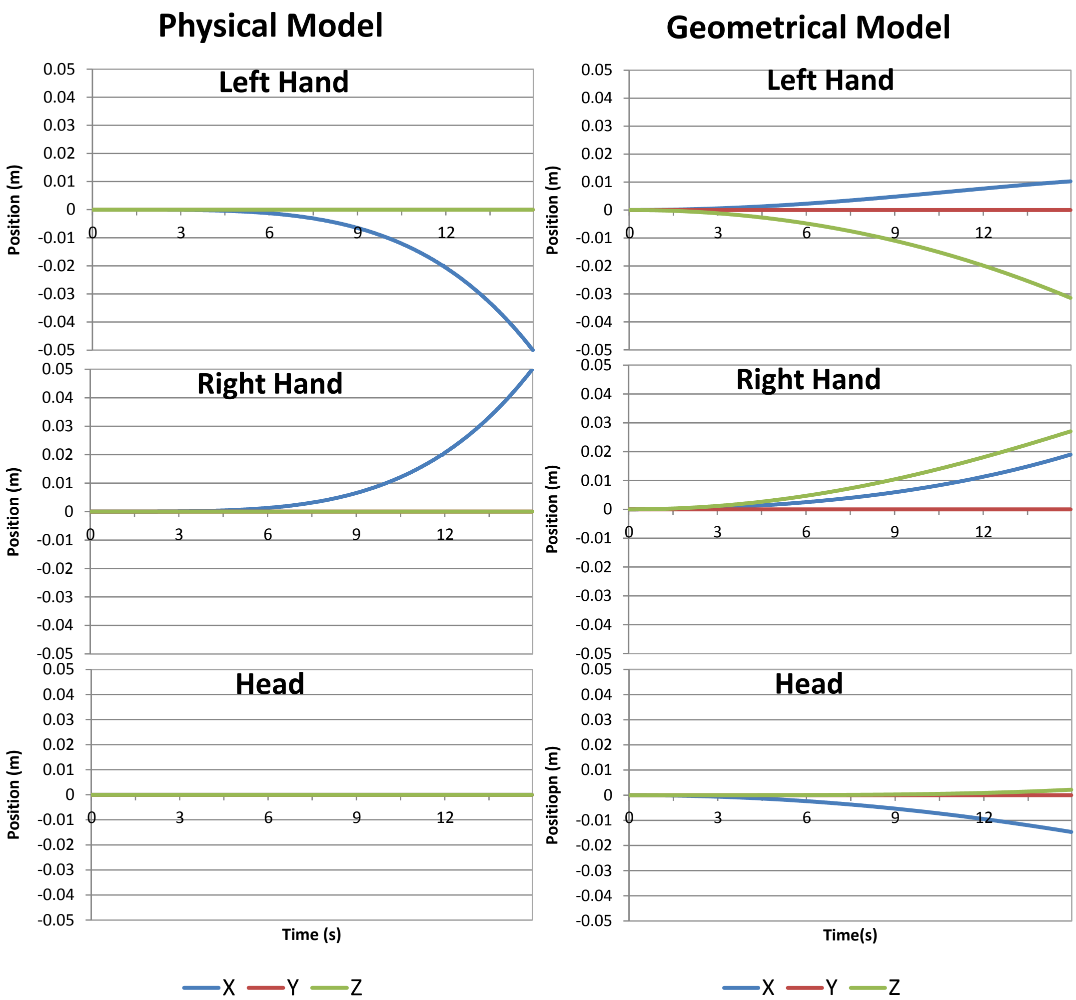


Figure 3.3: Output of the models. (*Physical* on the left, *Geometrical* on the right) for a left rotation around Y-axis of 15 seconds. The position in meters is plotted for each actuator LA, RA and H, and for each axis.

3.3.4 Haptic Rendering

Whatever the model selected to control the chair, for each instant t of the simulation, each actuator A (namely H, LA and RA) has to be in its targeted position G'_A (namely G'_H, G'_{LA} and G'_{RA}).

Most force-feedback devices (such as Novint Falcons) are impedance haptic devices, and the position of the actuator is thus not directly controllable. Indeed this kind of device is designed to sense the current position of the actuator and to provide a reaction force to the user. A classical spring-damper model may be used to control these devices in pseudo-position. The force F_A applied to an actuator A is computed by:

$$F_A = k(G'_A - P_A) - dV_A \tag{3.8}$$

where G'_A is the targeted position, P_A the current position of the actuator, V_A its velocity, k the spring constant and d the damping constant.

The models and the rendering algorithm were integrated to a home-made multimedia player that allowed the haptic rendering on three force-feedback devices to be synchronized with the audiovisual playback. The haptic loop runs at 1KHz and the value of the force F_A is updated at each instant t. The control software is written in C++ and runs on an ordinary personal computer.

3.4 User study

A user study was conducted to evaluate the quality of the simulated motion and to quantify its impact on the user's perceived quality of experience (QoE).

Seventeen participants took part in the study, aged from 21 to 54 ($\bar{x}$=36.11 σ_x=11.11). Five were female, two participants were left-handed and one already used a force-feedback device. The pilot study was presented as a single experiment lasting 20 to 30 minutes. Each participant was first introduced to the Novint Falcon and given a demonstration of its force capabilities. This step aimed to reduce the "surprise effect" for novice users. Participants were asked to passively experience each stimulus (see Figure 3.4 and Section 3.4.1) and then answer a questionnaire (see Section 3.4.3).

Figure 3.4: The user, comfortably installed on our device, is experiencing passive navigation enhanced by a haptic effect of motion.

3.4.1 Sequences: haptic-audiovisual contents

Two driving sequences were created to test our device, and evaluate the sensation of motion and quality of experience. We generated two 1-minute videos and the associated descriptions of the global motion in terms of $a(t)$ and $w(t)$. Our first sequence was a video of a **real car** driving session (see Figure 3.5a). Data was first captured using a front passenger equipped with a camera and an inertial measurement unit that sampled data at 30Hz (see Chapter 5 for more details on the capture setup).

The second sequence was a video of a **virtual car** racing video game (see Figure 3.5b). The main camera of the 3D simulation was placed inside the car in order to have a passenger point of view of the race. The visual output of the simulation was recorded while the accelerations and turn-rates of the car were extracted at 50Hz from the physics engine.

(a) Real video sequence of a car driving. (b) Virtual car race.

Figure 3.5: Haptic-Audiovisual contents.

3.4.2 Variables

To evaluate the quality of the simulated motion (and of the models) and the impact of this haptic feedback on the QoE, we defined four types of haptic feedback to be rendered with each sequence. **Physical Feedback**, computed from the physical model; **Geometrical Feedback** derived from the geometrical model; **No Haptic Feedback** in which only the audiovisual content was displayed, serving as a control to show how the other conditions impact on the QoE; and lastly **Random Haptic feedback** was provided. This random feedback was derived from low-pass filtered white noise (cutoff frequency $F_c = 0.5Hz$) played throughout the video. The amplitude of the signal was limited to the capabilities of the actuators. This last feedback was not synchronized with the video and was used to evaluate the effect of providing a continued haptic feedback.

All **height conditions** (two videos sequences $\times$ four types of haptic feedback) were presented in random order to the participants. They were not aware of the different types of haptic feedback.

3.4.3 Measurement of QoE: questionnaire

A questionnaire was designed to evaluate the QoE of passive navigation enriched with haptic feedback. QoE relates to the subjective user experience with a service or an application [Jai04, Kil08]. In our context this may be specified as the measure of the user's subjective experience with haptic-audiovisual content. In order to evaluate this experience, we built the questionnaire around the **Usability** [TA08] and the **Presence** [WS98] concepts.

Usability is defined by the norm ISO 9241-11 and aims at measuring how easy a system is to use. Three factors composed this concept: Efficiency, Effectiveness and Satisfaction. This latter measures how well the user enjoyed the system. “Effectiveness” means how well a user can perform a task while “Efficiency” indicates how much efforts are required. These two factors were not totally suitable for our system in the sense that it was not designed to perform a task. We preferred to use the term of *Comfort* to measure how well was the system to provide feedback. *Satisfaction* was however fully relevant in our situation.

Presence aims at measuring how much the user feels being physically situated in a virtual environment. Witmer and Singer [WS98] have identified four factors to determine the presence: Control, Sensory, Realism and Distraction. “Control” determines how much the user can control and modify objects within the virtual environment. “Sensory” characterizes how each sensory modality is solicited during the interaction. “Realism” describes how much the environment is realistic and consistent with user’s representation of the real world. “Distraction” identifies how much the user is disturbed by the apparatus used to create the virtual world. From this definition we focused on two factors: *Realism* and *Sensory*. As the user is passive with our system, Control factor was not relevant here. The Distraction factor was not directly used but included in the Comfort factor described previously.

The questionnaire was thus based on the four factors we wanted to evaluate: Realism, Comfort, Sensory and Satisfaction. Each factor was evaluated by questions rated on a 5-point scale. A mean was calculated for each factor. The sum of the scores gave a global QoE score. Table 3.1 presents the questions used to evaluate the QoE.

Factor	**Question**
Realism	How much did this experience seem consistent with your real-world experiences?
	How strong was your feeling of self-motion?
Sensory	How much did the haptic feedback contribute to the immersion?
	Were the haptic and visual feedback synchronized together?
Comfort	Was the system comfortable?
	How distracting was the control mechanism?
Satisfaction	How much did you enjoy using the system?

Table 3.1: QoE Questionnaire. Each question is rated on a 5-point scale from 1 (Not at all) to 5 (Totally)

3.4.4 Results

Two hypotheses are tested: the *HapSeat* enhances the quality of experience, and it does generate a sensation of motion. Shapiro-Wilk and Bartlett tests were performed on our data and the normality and homoscedasticity for most distributions could not be assumed. Hence non-parametric tests were used to analyze the results presented in this section.

As described above, a score for the four factors, Realism, Sensory, Comfort and Satisfaction were obtained using a questionnaire (see Figure 3.6 and Table 3.2). First, the main result confirms our first hypothesis. Our device significantly enhances the quality of experience (Friedman Anova: $p = 8.44e^{-08} < 0.05$). The Wilcoxon test with the Holm-Bonferroni correction has been used for the post-hoc analysis (see Table 3.3). With the haptic feedback computed from the *Physical* or *Geometrical* model, the QoE is significantly higher than without haptic feedback ($p < 0.05$). However the QoE of the *Geometrical Model* is not significantly different from the QoE of the *Physical Model* ($QoE_G = 15 \approx QoE_P = 14.20, p = 0.5575 > 0.05$). Second, it seems that haptic feedback consistent with the video is necessary to improve the QoE: user scores for random feedback are not statistically different to the no feedback condition ($QoE_N = 8.36 \approx QoE_R = 9.45, p = 0.4816 > 0.05$).

This tendency is observable for three factors. Presenting users with haptic feedback computed from our models resulted in significant increase in their reporting of Realism (Friedman Anova, $p = 3.80e^{-08} < 0.05$), Sensory (Friedman Anova, $p = 7.02e^{-08} < 0.05$) and Satisfaction (Friedman Anova, $p = 3.86e^{-07} < 0.05$) scores. However Comfort remained similar for all conditions: the Friedman Anova is significant, $p = 1.27e^{-03} < 0.05$, but Wilcoxon tests cannot confirm this hypothesis, $p > 0.05$ (see Table 3.3).

Finally no significant differences are found for the QoE of each model between the two sequences *Real Car* and *Virtual Car* (Wilcoxon test, $p_{Geo} = 0.3933$ and $p_{Phy} = 0.4173 > 0.05$).

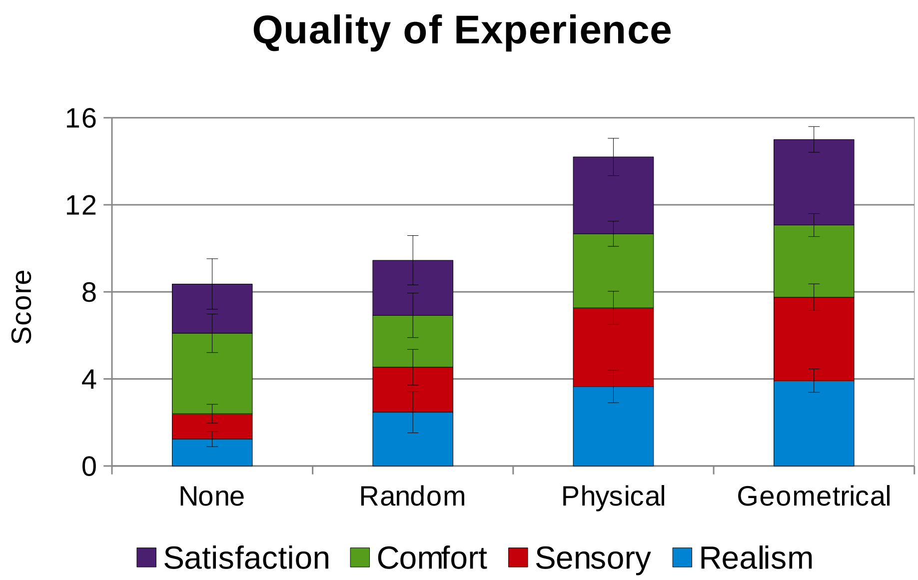


Figure 3.6: Quality of experience. The haptic feedback provided by the Physical and Geometrical models significantly improves the QoE.

In order to evaluate the second hypothesis, which is that the *HapSeat* generates

Factor **Model**	QoE	Realism	Sensory	Comfort	Satisfaction	
None	8.3578	1.2353	1.1618	3.6961	2.2647	$\bar{x}$
	2.0741	0.3477	0.4325	0.8853	1.1608	σ_x
Random	9.4479	2.4688	2.0625	2.3854	2.5313	$\bar{x}$
	2.9550	0.9481	0.8190	1.0187	1.1324	σ_x
Physical	14.1961	3.6471	3.6176	3.4020	3.5294	$\bar{x}$
	2.5521	0.7451	0.7609	0.5790	0.8564	σ_x
Geo.	15	3.9167	3.8333	3.3166	3.9333	$\bar{x}$
	1.7904	0.5401	0.6099	0.5300	0.5936	σ_x
F. Anova	35.7534	37.3958	36.1324	15.7554	32.6279	χ^2
	3	3	3	3	3	df
	$8.44e^{-8}$	$3.80e^{-8}$	$7.02e^{-8}$	$1.27e^{-3}$	$3.86e^{-7}$	p
	***	***	***	*	***	sig.

Table 3.2: Means ($\bar{x}$) and Standard deviations (σ_x) for each model with respects to each factor. A Friedman Anova ($\chi^2, df, p.value$) has been performed on each factor.

QoE	Geometrical	None	Physical
None	$1.5e^{-05}$	-	-
Physical	0.5575	$6.5e^{-05}$	-
Random	$6.9e^{-05}$	0.4816	0.005

Realism	Geometrical	None	Physical
None	$5.3e^{-06}$	-	-
Physical	0.4336	$4.1e^{-06}$	-
Random	0.0005	0.0004	0.0028

Sensory	Geometrical	None	Physical
None	$5.5e^{-06}$	-	-
Physical	0.5169	$5.2e^{-06}$	-
Random	$4.6e^{-05}$	0.0004	0.0002

Comfort	Geometrical	None	Physical
None	0.1575	-	-
Physical	0.4927	0.1664	-
Random	0.0161	0.0064	0.0107

Satisfaction	Geometrical	None	Physical
None	0.002	-	-
Physical	0.4992	0.0095	-
Random	0.0037	0.4992	0.0273

Table 3.3: Pairwise comparison of each model for each factor using Wilcoxon test with Holm-Bonferroni correction.

a sensation of motion, the answers to the two questions of the Realism factor were analyzed (see Figure 3.7, Q1 on top and Q2 on bottom and Table 3.4). The results from Q1 suggest that the simulated motion was perceived as realistic (Friedman Anova $p = 3.60e^{-05} < 0.05$). A Wilcoxon test with the Holm-Bonferroni correction was also performed on our data (see Table 3.5). Again, no statistical difference between the *Physical* and *Geometrical* models is observed ($Q1_P = 3.6 \approx Q1_G = 3.8$, $p = 0.6356 > 0.05$) but they are significantly different from the *Random* and *None* conditions ($p < 0.05$). The results from Q2 follow the same pattern. Both models provided a strong sensation of motion, significantly higher than *Random* and *None* conditions (Friedman Anova $p < 0.05$, Wilcoxon tests $p < 0.05$).

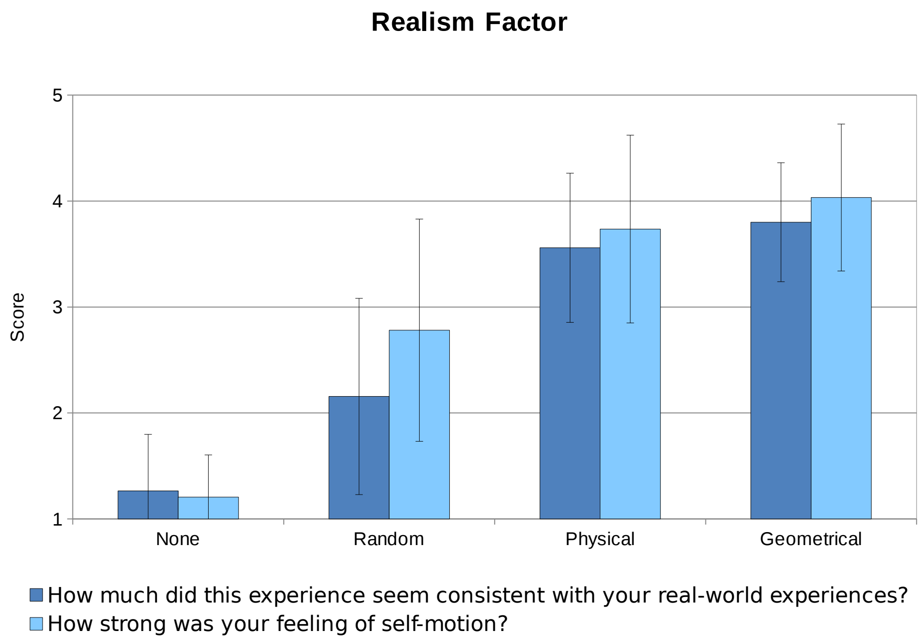


Figure 3.7: Realism factor details. Users found the simulation realistic and experienced a strong sensation of self-motion.

Question / **Model**	Q1	Q2	
None	1.2647	1.2059	$\bar{x}$
	0.5338	0.3976	σ_x
Random	2.1563	2.7813	$\bar{x}$
	0.9259	1.0483	σ_x
Physical	3.5588	3.7353	$\bar{x}$
	0.7045	0.8860	σ_x
Geometrical	3.8	4.0333	$\bar{x}$
	0.5606	0.6935	σ_x
F. Anova	20.4615	12.86	χ^2
	2	2	df
	3.60^{-05}	$1.61e^{-03}$	p
	***	*	sig.

Table 3.4: Means ($\bar{x}$) and Standard deviations (σ_x) for each model with respects to Q1 and Q2. A Friedman Anova ($\chi^2, df, p.value$) has been performed on each question.

Q1	Geometrical	None	Physical
None	$4.5e^{-06}$	-	-
Physical	0.6356	$3.9e^{-06}$	-
Random	0.0002	0.0030	0.0005

Q2	Geometrical	None	Physical
None	$3.5e^{-06}$	-	-
Physical	0.3743	$3.5e^{-06}$	-
Random	0.0045	0.0001	0.0238

Table 3.5: Pairwise comparison of each model for both question using Wilcoxon tests with the Holm-Bonferroni correction.

3.4.5 Discussion

Our results suggest that the *HapSeat* does enhance the user experience during passive navigation simulation. Both rendering models significantly increased the QoE compared to the Random and None feedback conditions. Our results also suggest that the synchronization of the haptic effect with the visual content is important.

In this study, no statistical differences are found between the *Physical* and *Geometrical* models. This is probably due to the nature of the simulation (car driving) which does not fully exploit the 6DoF. Only translation (car acceleration) and rotation (turns) were included in the two sequences tested. More complex content, such as spaceship flight or a rollercoaster ride, might produce results that highlight differences between the models. In addition, the parameters of the models could be tuned to increase their differences. Each one is composed of several factors which impact the use of the workspace. The *Physical Model* could also be improved by modeling the segments and joints of the user's skeleton instead of treating the user as a single rigid body.

We observed that the simulated motion was not perceived in the same way by all participants. Some of them found that the haptic feedback computed from our models was reversed. For instance, they expected to be pushed backward instead of being pulled forward when the car (real or virtual) was moving straight forward. However this observation was not consistent among all users. Some participants expected to feel the reaction force instead of the acceleration only during turns, but found the feedback acceptable for linear translations, i.e. when the car was going straight forward. Though some participants seem to prefer a reversed force feedback in specific cases, this does not mean that the output of the models should necessarily be reversed. One might posit two user profiles "direct" and "reversed" to address this, it can certainly be said that the design of the associated haptic feedback is not straightforward. The perception of motion simulated by force-feedback devices requires further evaluation. Studies are also needed to understand the influence of a haptic stimulus on the perception of a visual stimulus.

Our device was reported as comfortable and user friendly. The perception of comfort was similar with and without haptic feedback, suggesting that no extra discomfort is introduced by the system. Nevertheless comfort could be improved, especially for the headrest. Some participants reported that the haptic feedback for the *real car* sequence contained too much vibration. This may be explained by the greater sensitivity of proximal joints to movement than distal joints [Jon00]. Similar displacements are perceived more strongly on the head than on the hands. If this vibration that contribute to realism when perceived by the user's hand, might be too intense for the head. So far the haptic rendering for all actuators is the same. But dedicated algorithms could be implemented for each device. As a minimum, a low-pass filter could be applied on the output of the actuator H to reduce vibration. Attenuation coefficients can also be added and adjusted depending on the preference of the end-user.

3.5 Chapter conclusion

In this chapter, we have introduced the *HapSeat*, a novel approach to the simulation of 6DoF effect of motion. Instead of moving the whole user's body as it is traditionally done with motion platforms, we stimulate only parts of the body. Our hypothesis was that, coupled with a visual stimulus, these local stimulations could trigger a sensation of motion and thus improve the quality of experience.

We used three force-feedback devices to stimulate the user's hands and head when seated. A proof-of-concept prototype has been built, which rely on these three devices to simulate two moving armrests and a moving headrest. Two models were implemented to generate the effects of motion. The *Physical* model computes the forces supposed to be felt during a movement. The *Geometrical* model modifies the structure of the chair to match the position that characterizes the movement.

A user study has been conducted to validate our concept. A methodology and metrics have been designed for this purpose. Results of the study show that the two control models succeed at enhancing the quality of experience during passive navigation. Several factors have been identified to measure the quality of experience (Realism, Comfort, Sensory, Satisfaction), and a dedicated questionnaire was designed. Furthermore participants reported having experienced a realistic sensation of self-motion. Thus it seems that our approach yields a new way to simulate a sensation of motion in a consumer environment and allows the creation of more immersive applications.

Chapter 4

Haptic rendering for HAV based on a washout filter

Contents

In the context of haptic-audiovisuals (HAV), recent works have proposed graphical authoring tools to create and synchronize haptic feedback to audiovisual content [WRTH13, Kim13]. These new tools allow to easily design haptic effects without knowledge on the control of haptic devices. But they also bring new challenges for the rendering of haptic effects.

Haptic effects are designed independently from a specific device and its workspace. The haptic rendering algorithm has then to adapt the effects to the constraints of the device. Besides multiples haptic effects may have been created. The transitions between these effects must also be handled by the haptic rendering.

In this chapter we propose to improve the haptic rendering of HAV with a washout filter based on the human kinesthetic perception. A body model and an inverse kinematics algorithm are used to determine the user's kinesthetic thresholds. The concept of our washout filter is introduced in section 4.1. The implementation is detailed in section 4.2, followed by the user study in section 4.3. Results are discussed in section 4.4 and the application of a full video sequence is presented in section 4.5. Finally conclusions are provided in section 4.6.

4.1 Washout filter for haptic rendering

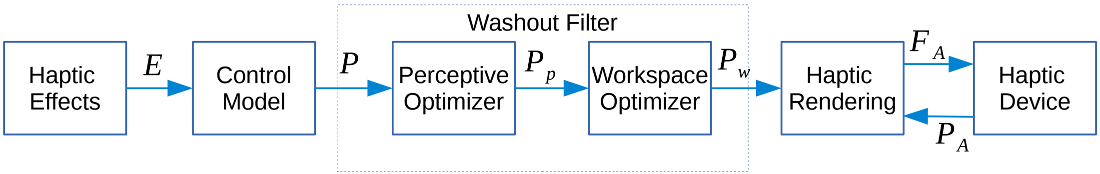


Figure 4.1: Washout filter for haptic rendering. Haptic effects E are converted into positions P for the actuator by a control model. The perceptive optimizer removes undesirable effects by limiting the movement of the actuator (P_p), and the workspace optimizer ensures that the limits of the workspace are respected (P_w). The force F_A is then rendered by the device depending on its current position P_A.

Content creators can easily add trajectories or motion effects to audiovisual content thanks to authoring tools [WRTH13, Kim13]. For example, three separate forward movements may be defined (see Figure 4.2). Then a control model adapts the effects to the capabilities of the actuator and three forward movements are rendered (see Chapter 3 for an example of a control model). But, at the end of an effect, the actuator suddenly goes back to the central position, inducing a "counter-effect" which should not be perceived by the user.

This problem has been solved for motion platforms where a *washout filter* is in charge of tacking back the device to a neutral position without making the user aware of this process. The platform is moved under the perceptual threshold of the vestibular system which is responsible for the sensation of motion (around $0.1 m.s^{-2}$ [HJZ+02]). However these algorithms are not suitable for force-feedback devices which do not stimulate the vestibular system, but the kinesthetic system (perception of the force and movements).

To tackle this issue, we propose a new workflow for the haptic rendering of HAV, based on a washout filter (see Figure 4.1). This washout filter reduces these counter-effects while preserving the actual effects. Two steps compose the washout filter. First, the perceptive optimizer relies on kinesthetic perceptual thresholds to make the counter-effects imperceptible. Then the workspace optimizer ensures that the maximum space of the workspace is exploited and that its limits are respected. Figure 4.2 shows the results of the washout filter on the three effects of our example.

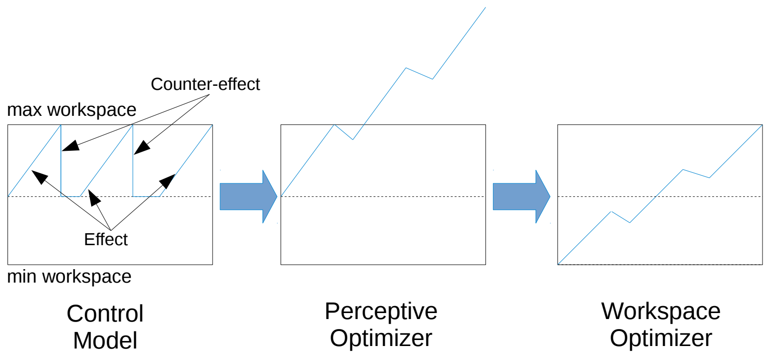


Figure 4.2: Schematic example of the use of our washout filter. Three effects have been designed. Between the effects, the actuator suddenly goes back to the central position (counter-effects). The washout filter removes these counter-effects (Perceptive Optimizer) and optimizes the use of the workspace (Workspace Optimizer).

4.2 Proof-of-concept

Our proof-of-concept is designed with the *HapSeat*, a force-feedback based motion simulator (see Chapter 3). The haptic effects are first formalized in this section, followed by a description of the device and the associated control model. Then the implementation of the washout filter is described and its performances are evaluated.

4.2.1 Haptic effects

Haptic effects were designed with an editor allowing to synchronize them with a video (see Chapter 5). A haptic effect E_i of a global set of N haptic effects $E = \{E_i\}_{1 \leq i \leq N}$ starts at an instant T_i of the video, has a duration of D_i, and is described by M_i, a force (Cartesian coordinates x_c, y_c, z_c) and a torque (three Euler angles ϕ_c, θ_c, ψ_c) at an instant $t \in [T_i, T_i + D_i]$:

$$M_i(t) = [x_c^i(t), y_c^i(t), z_c^i(t), \phi_c^i(t), \theta_c^i(t), \psi_c^i(t)] \tag{4.1}$$

4.2.2 Haptic device: the HapSeat

The *HapSeat* simulates motion sensations in consumer settings using local force-feedback devices (see Chapter 3). In the remainder of this paper the following notation is used. The actuators near to the head, left hand and right hand are labelled H, LA, and RA. Their central positions in their workspaces are named respectively G_H, G_{LA} and G_{RA}, G being the center of the space. The workspace of one actuator is defined by W (in our case the dimension of W is 10×10×10cm).

4.2.3 Control model

For a given haptic effect E_i, the control model for one local actuator A is formulated in terms of displacement from its initial and central position G_A to the new position

G'_A, $\forall t \in [T_i, T_i + D_i]$:

$$\overrightarrow{G_A G'_A}(t) = f(\vec{T}(t), \vec{R}(t)) \tag{4.2}$$

where

$$f(\vec{T}(t), \vec{R}(t)) = \frac{\|\vec{T}(t)\|\vec{T}(t) + \|\vec{R}(t)\|\vec{R}(t)}{\|\vec{T}(t)\| + \|\vec{R}(t)\|} \tag{4.3}$$

and

$$\vec{T}(t) = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{bmatrix} \begin{bmatrix} x_c^i(t) \\ y_c^i(t) \\ z_c^i(t) \end{bmatrix} \tag{4.4}$$

$$\vec{R}(t) = (R_x(m_x \phi_c^i(t)) R_y(m_y \theta_c^i(t)) R_z(m_z \psi_c^i(t)) - I_3)\overrightarrow{GG_A}) \tag{4.5}$$

The function f is the combination of two vectors $\vec{T}$ and $\vec{R}$ which respectively uses the positions and rotations described by the trajectory M_i. The scaling factors s_x, s_y, s_z, m_x, m_y, m_z map the motion effect to the workspace of the actuator. R_x, R_y and R_z are the 3D rotation matrices around their respective X, Y and Z axes and I_3 is the identity matrix in dimension 3.

From this equation, the new application points G'_H, G'_{LA} and G'_{RA} are computed from the initial points G_H, G_{LA} and G_{RA}. The scaling factors are computed to use the workspace of each actuator in an optimal way, i.e. avoiding any saturation while using the largest space available. The computation of those scaling factors is performed by a preprocessing step consisting in finding the maximal amplitude of the displacement rendered by the three different actuators (H, LA, and RA).

Eventually the position P computed by the control model for one actuator A can be defined as:

$$P(t) = \begin{cases} \overrightarrow{G_A G'_A}(t), \forall t \in [T_i, T_i + D_i] \\ 0, \forall t \in]T_i + D_i, T_{i+1}[\end{cases} \tag{4.6}$$

Between haptic effects ($\forall t \in]T_i + D_i, T_{i+1}[$), the actuator goes to the central position, which may induce counter-effects. Hence we propose the following washout filter to remove them.

4.2.4 Washout filter

4.2.4.1 User's body model

The kinesthetic perception is complex and includes several factors. According to Jones [Jon00], the kinesthetic perception is related to the angular speed of the joints of the moving limb. The faster the movement is, the lower the detection threshold is. Besides the joints do not have the same sensibility: proximal joints are more sensitive than distal ones. Muscles play also a role in the detection threshold which decreases with the contraction of the muscles.

Our approach consisted in designing a user's body model to compute the kinesthetic perception. In this model we assume that the user is relaxed when watching a video, then the user's perceptual thresholds are determined by angular speeds associated to each joint. The arms are considered as two segments (arm and forearm) and two joints (elbow and shoulder). The head and neck are composed by one segment and one joint respectively. The size of segments and the angle limits of joints are defined by anatomical data [SSS+10] and listed in Table 4.1. Constraints for the shoulders are not necessary because the movements imposed by the actuators are too small.

Limb	**Size (cm)**
Neck and Head	16.8
Forearm	36.2
Arm	34.9
Positions	**x,y,z (cm)**
Left Shoulder	(30,30,0)
Right Shoulder	(-30,30,0)
Base Neck	(0,30,0)
Left Actuator (G_{LA})	(30, -10, 40)
Right Actuator (G_{RA})	(-30, -10, 40)
Head Actuator (G_{HA})	(0, 46.8, 0)
Constraints	**Angle (deg.)**
Elbow	0 to 140
Head (pitch)	-65 to 40
Head (yaw)	-50 to 50
Head (roll)	-35 to 35

Table 4.1: Biomechanical constraints and parameters used in our model.

4.2.4.2 Perceptive optimizer

The perceptive optimizer consists in moving the actuator toward its central position in a way that this movement is not perceived by the user. Hence the actuator speed needs to stay under the user's perceptual threshold.

Determining the angular speed of joints is a well known problem in computer animation. Given the skeleton of an arm composed by segments and joints and the given target position of the hand, multiple combination of angular speeds are possible. This kind of problem is solved by inverse kinematics algorithms. We use here the "cyclic coordinate descent" (CCD) algorithm which is effective for simple configurations with small amplitude of movements [Wel93]. The CCD is an iterative method which minimizes the distance between the end effector (in our case the hand or head) and the target position (the actuator) by modifying the angle of each joint. Starting from the joint closest to the end effector the algorithm iterates through the kinematic chain to the farthest.

The new position of the actuator computed by the perceptive optimizer can be defined as:

$$P_p(t) = \begin{cases} P_p(T_i) + P(t), \forall t \in [T_i, T_i + D_i] \\ P_p(t - \Delta t) + v\Delta t, \forall t \in]T_i + D_i, T_{i+1}[\end{cases} \tag{4.7}$$

with Δt the sampling time (typically 20ms) and

$$v = \underset{0 \leq K(v,t)}{\arg\min}(\|P(t - dt) + v\Delta t\|) \text{ and } P_p(0) = 0 \tag{4.8}$$

and the function K defined in Figure 4.3

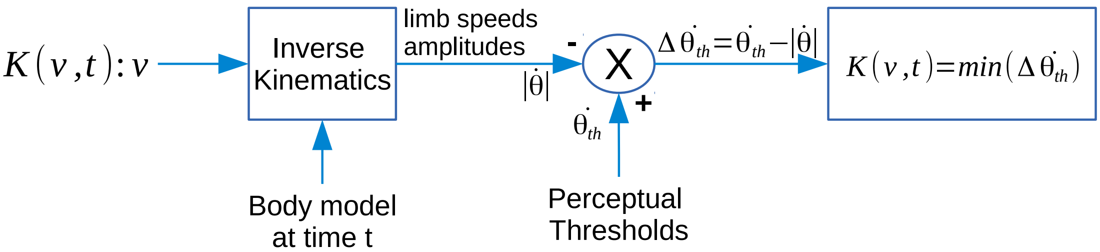


Figure 4.3: Function K determining the speeds of the actuator. The angular speeds of the joints are computed from the current speed v of the actuator. The speeds constrained by the perceptual thresholds are then calculated.

P_p can also be written:

$$P_p(t) = [P_p^1(t), P_p^2(t), P_p^3(t)]^T \tag{4.9}$$

where P_p^1, P_p^2, P_p^3 are the positions of P_p for the axes X, Y and Z respectively.

4.2.4.3 Workspace optimizer

The previous step does not guarantee that the positions computed respect the limits of the workspace (see Figure 4.2). The workspace optimizer makes sure that these positions are compatible with the workspace of the actuator. An offset is performed to use the maximum space, and if necessary, the amplitude of the haptic effects is reduced. The new position computed by the workspace optimizer can be formalized as:

$$P_w(t) = \begin{cases} P_p(t) \text{ if } \forall t, P_p(t) \in W \\ \dfrac{P_p(t) - O_p}{s^*} \text{ else} \end{cases} \tag{4.10}$$

with

$$s^* = \min(s | \forall t, \frac{P_p(t) - O_p}{s} \in W) \tag{4.11}$$

where $s \geq 1$ and

$$O_p = [O_p^1, O_p^2, O_p^3]^T \tag{4.12}$$

where

$$O_p^i = \left\{ \frac{\max_t(P_p^i(t)) - \min_t(P_p^i(t))}{2} \right\}_{1 \leq i \leq 3} \tag{4.13}$$

4.2.5 Haptic rendering

The washout filter provides, at each instant t, the target position P_w for each actuator A. Most force-feedback devices (such as the Novint Falcons) are impedance haptic devices, and the position of the actuator is thus not directly controllable. Indeed this kind of device is designed to sense the current position of the actuator and to provide a force feedback to the user. The actual haptic rendering is performed thanks to a spring-damper model. The force $\vec{F_A}$ applied to an actuator A is computed by:

$$\vec{F_A} = k(\vec{P_w} - \vec{P_A}) - d\vec{V_A} \tag{4.14}$$

where $\vec{P_w}$ is the targeted position, $\vec{P_A}$ the current position of the actuator, $\vec{V_A}$ its velocity, k the spring constant and d the damping constant.

4.2.6 Implementation

4.2.6.1 Perception thresholds

The perception thresholds are key values in our system because they determine how much the counter-effects are reduced. These effects must be imperceptible. We defined thresholds by referring to the results described by Jones [Jon00]. The perceptual threshold for the elbow is around 1 $deg.s^{-1}$ (angular speed of the joint). Proximal joints are known to be more sensitive than distal joints, thus we could set the thresholds for the shoulder and the neck to 0.5 $deg.s^{-1}$.

Three profiles of perception thresholds have been designed (see Table 4.2). First one, entitled T1, is based on Jones' results. Such angular speeds are small, and to respect these constraints the actuators have to move very slowly. The positions P_p computed might probably not fit the workspace (see Figure 4.2). The scaling factor s^* applied by the workspace optimizer might strongly reduce the actual effects. Thus we also propose less restrictive constraints to preserve the amplitude of the effects with the profiles T2 and T3. Besides in a context of HAV, the user's attention is split between haptic and audiovisual feedback. Movements of an actuator might thus be not perceived with higher thresholds. Hence thresholds for T2 are higher than for T1, and those for T3 are higher than for T2. They were set empirically. As the output of the washout is non linear, determining such thresholds is not trivial. With T2 and T3 the counter-effects may be perceptible, but the scaling factor applied by the workspace optimizer will be lower than with the profile T1.

T1	**Threshold** $(deg.s^{-1})$
Neck	0.5
Shoulder	0.5
Elbow	1
T2	**Threshold** $(deg.s^{-1})$
Neck	1.5
Shoulder	1.5
Elbow	2.5
T3	**Threshold** $(deg.s^{-1})$
Neck	3
Shoulder	3
Elbow	4

Table 4.2: Three profiles determining the user's perceptual threshold.

4.2.6.2 Performance evaluation

A first set of tests was conducted to assess the properties of our washout filter for each of the three profiles.

We have first measured the time required for each profile to move an actuator from the edge of its workspace to the central position. The distance from the border to the center for the Novint Falcon is 5 cm. As expected, the lowest the perceptual threshold, the longest the time required to reach the central position: 1.8 secs for T3, 4.7 secs for T2 and more than 10 secs for T1 (see Figure 4.4).

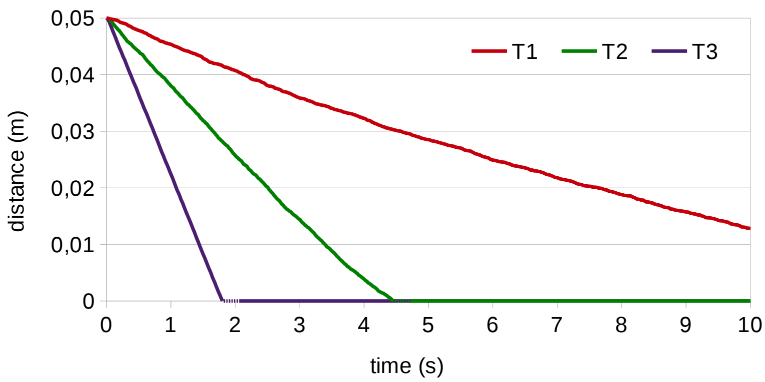


Figure 4.4: Time required for each profile, T1, T2 and T3, to move the actuator from the edge of its workspace to the central position.

Second, we have evaluated the impact of each profile on the amplitude of the filtered effects. We have created several 15-second sequences with two to seven successive effects (as described in Section 4.2.1). The effects are identical which is the most critical situation for the haptic rendering: the actuator has to go several times in the same direction. Examples are available in Appendix A with sequences composed by three, four and five effects (Figures A.1, A.2 and A.3 respectively). The more effects there are, the less time available there is for the perceptive optimizer, and the stronger the scaling s^* applied by the workspace optimizer (see Equation 4.11). Figure 4.5 depicts this scaling factor. We observed that with the profile T1, starting from two consecutive effects, the amplitude of the effects is decreased. As expected, the profile T2 less impacts the amplitude and the profile T3 even less. But when considering seven effects in a sequence the amplitude is reduced by at least 30% for each profile.

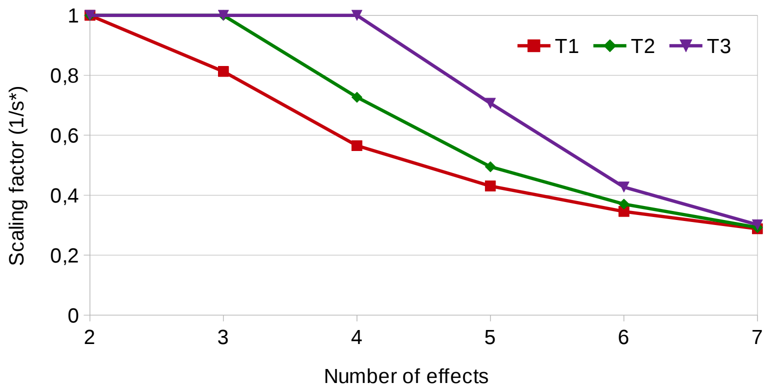


Figure 4.5: Scaling factor applied by the workspace optimizer. The more effects there are in a sequence, the more down-scaled they are.

To conclude on this section, we observe that the profile T1 removes the counter-effects, but it has a cost on the amplitude of the actual effects. By increasing the perceptual thresholds, with the profiles T2 and T3, amplitudes are more preserved but the counter-effects would become perceptible.

4.3 User study

A user study was conducted to evaluate the relevance of this new haptic rendering for HAV. Our hypothesis is that the quality of haptic-audiovisual experience is better when counter-effects are not perceived (i.e. in presence of our washout filter). We also wanted to explore the influence of the thresholds on the quality of experience (QoE). We therefore tested the three profiles (T1, T2 and T3) and analyzed the perception of the QoE by users.

Twenty participants took part in this experiment, aged from 23 to 52 ($\bar{x}$=39.7 σ_x=9.21). Five were female and three left-handed. None of them was an expert user of force-feedback devices or motion platforms.

4.3.1 Experimental conditions

To evaluate the impact of our washout filter on the haptic-audiovisual experience, we used four types of haptic feedback: **T1**, **T2** and **T3** respectively computed using the T1, T2 and T3 profiles. The fourth haptic feedback, **T0**, bypasses the washout filter.

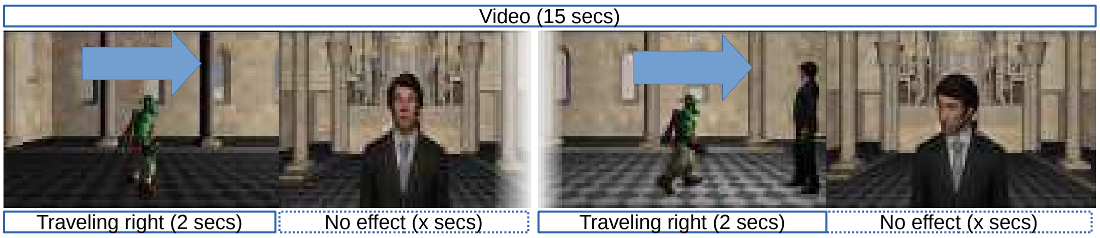


Figure 4.6: Structure of the 15-second video sequences used in the user study. A video is composed by a succession of two sequences: a two-second traveling following a green character, and an x-second still shot of a man.

The haptic rendering is directly computed from the positions P given by the control model. This feedback serves as a control condition to determine how the washout filter and the profiles modify the quality of experience.

These four haptic feedbacks were compared against each others for three video sequences of 15 seconds, named **S1**, **S2** and **S3**. They were designed according to the structure depicted in Figure 4.6: combinations of sequences of a traveling shots of a walking character (2 secs) and a still shot of a man (x secs). S1 is composed by a succession of three sequences (x=3 secs), S2 of four sequences (x=1.74 secs) and S3 of five sequences (x=1 sec). A haptic effect was associated to each traveling (movement toward the right). Then these three sequences exploit the different outputs observed in section 4.2.6.2 (see Figures A.1, A.2 and A.3). Then **4×3 haptic-audiovisual contents** had to be experienced by each participant.

4.3.2 Procedure

The experiment lasted around 20 minutes for each participant, comfortably installed on the *HapSeat*. The study was divided in three steps corresponding to the three sequences S1, S2 and S3. The steps were performed in a random order. For each, the participant had the possibility to experience the video and the four haptic feedbacks. They were asked to try each haptic-audiovisual content, as many times as needed, and to order the different haptic rendering from 1 (the best) to 4. Then they went to step 2 and 3 to experience the two others video sequences and associated haptic feedbacks. Finally an informal interview was conducted to collect more information about the user's experience.

4.3.3 Results

The normality of the distributions cannot be assumed according to the Shapiro-Wilk test. Hence results were analyzed using non-parametric tests: Friedman Anova and Wilcoxon test with Holm-Bonferroni correction.

The main result of our study is that the three haptic feedbacks provided by the washout filter are preferred to the classical haptic feedback (see Figure 4.7, F. Anova: $p < 0.05$). The three profiles are not statistically different (Wilcoxon test: $p > 0.05$).

The analysis of the ranking for each video sequence provides the same results: the washout filter clearly improves the user's experience, and there is still no difference between the profiles. A deeper analysis of the participants' ranking is then necessary.

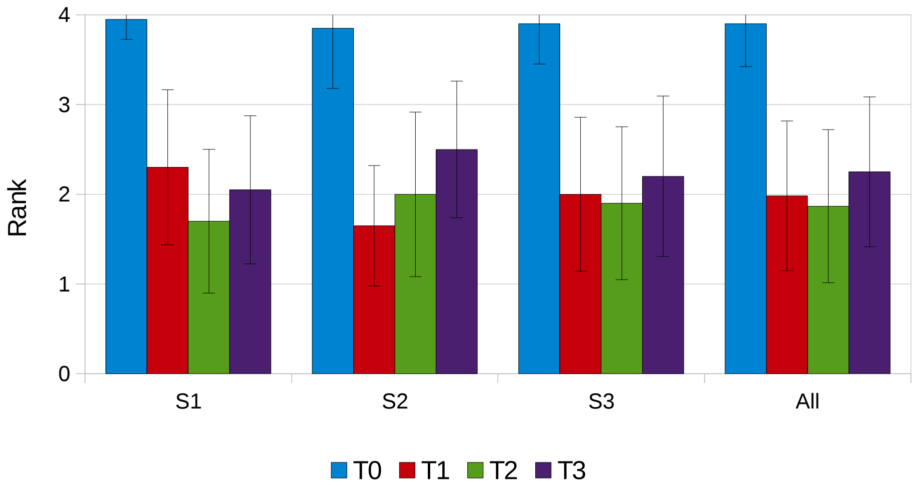


Figure 4.7: Rankings for all sequences. Conditions with a washout filter are preferred to the control condition (T0).

Sequence **Profile**	All	S1	S2	S3	
T0	3.9 0.48	3.95 0.22	3.85 0.67	3.9 0.44	$\bar{x}$ σ_x
T1	1.98 0.83	2.3 0.86	1.65 0.67	2 0.86	$\bar{x}$ σ_x
T2	1.87 0.85	1.7 0.80	2 0.92	1.9 0.85	$\bar{x}$ σ_x
T3	2.25 0.84	2.05 0.83	2.5 0.76	2.2 0.89	$\bar{x}$ σ_x
F. Anova	80.2471 3 $2.2e^{-16}$ ***	30.5294 3 $1.068e^{-6}$ ***	26.8588 3 $6.303e^{-6}$ ***	26.6471 3 $6.981e^{-6}$ ***	χ^2 df p sig.

Table 4.3: Means ($\bar{x}$) and Standard deviations (σ_x) for each profile with respects to each sequence. A Friedman Anova ($\chi^2, df, p.value$) has been performed on each sequence.

The informal interviews at the end of the experiment led to interesting observations. We roughly identified three groups of participants. Some of them preferred a perfect match between the haptic feedback and the video, when counter-effects were imperceptible. Some preferred more dynamic effects and were more tolerant regarding the perception of counter-effects. So they could better understand when an effect starts and stops. Finally some participants did not have any preference between the three profiles, the results were acceptable in any case. We performed then a hierarchical

All	T3	T2	T0
T2	0.071	-	-
T0	$2.0e^{-16}$	$2.0e^{-16}$	-
T1	0.144	0.568	$2.0e^{-16}$
S2	T3	T2	T0
T2	0.278	-	-
T0	$1.0e^{-05}$	$2.6e^{-05}$	-
T1	0.023	0.265	$7.8e^{-06}$

S1	T3	T2	T0
T2	0.18	-	-
T0	$1.1e^{-06}$	$9.0e^{-07}$	-
T1	0.57	0.12	$4.0e^{-06}$
S3	T3	T2	T0
T2	0.74	-	-
T0	$4.3e^{-06}$	$3.2e^{-06}$	-
T1	0.74	0.74	$7.3e^{-06}$

Table 4.4: Pairwise comparison of each sequence for each profile using Wilcoxon test with Holm-Bonferroni correction.

cluster analysis of the participants' ranking to determine if these three groups could be found [JMF99]. The distance between each participant was computed by an Euclidean distance on the whole set of rankings, and Ward's method was used for the clustering (clusters are computed in a way that their variance is minimal). Results are displayed on a dendrogram where three groups may be identified (cut at height = 8, Figure 4.8).

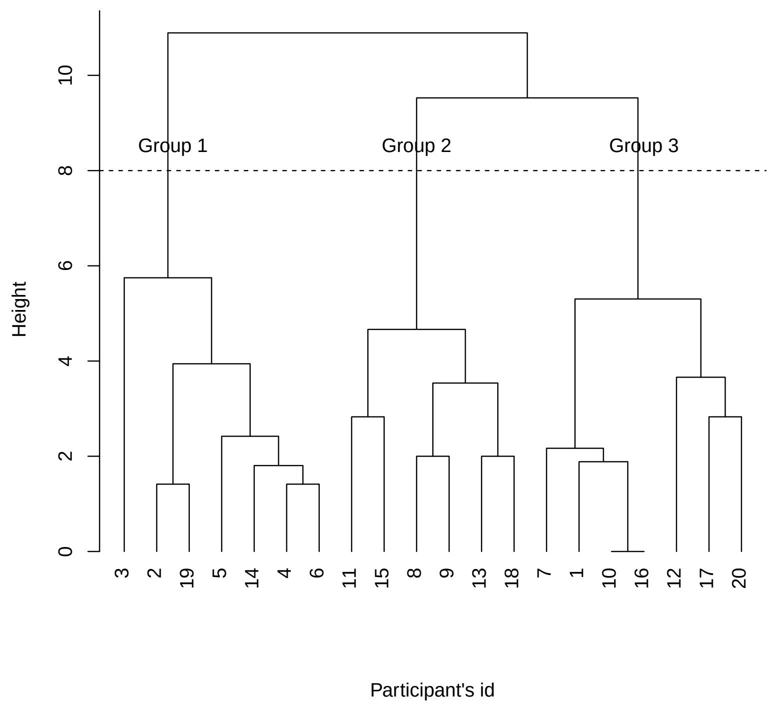


Figure 4.8: Dendrogram of the cluster analysis. Three groups of 6 or 7 participants emerge.

Our observations are confirmed by the analysis of the results of each group. The average ranking for Group 1, from the most preferred to the less, is: T1, T2, T3 and T0 (Figure 4.9). The four conditions are statistically different (F. Anova: $p < 0.05$, Wilcoxon $p < 0.05$). Group 3 has preferred the T2 profile (Wilcoxon test: $p < 0.05$), while T1 and T3 are not statistically different. Finally Group 2 reproduces the general results. The three profiles performed better than the T0 but they are not statistically different. Ranking for each sequence was also analyzed but they follow the patterns observed for each group.

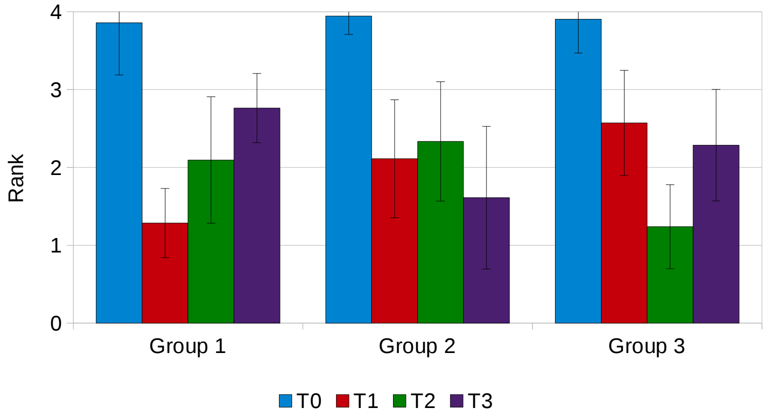


Figure 4.9: Ranking of the groups for all sequences. Group 1 prefers the profile T1 while Group 3 prefers T2. The three profiles are not statistically different for Group 2.

Sequence Profile	Group 1	Group 2	Group 3	
T0	3.86 0.67	3.94 0.24	3.90 0.44	$\bar{x}$ σ_x
T1	1.29 0.44	2.11 0.76	2.57 0.68	$\bar{x}$ σ_x
T2	2.10 0.81	2.33 0.77	1.24 0.54	$\bar{x}$ σ_x
T3	2.76 0.44	1.61 0.92	2.29 0.72	$\bar{x}$ σ_x
F. Anova	44.71 3 $1.1e^{-09}$ ***	33 3 $3.2e^{-07}$ ***	45.57 3 $7.0e^{-10}$ ***	χ^2 df p sig.

Table 4.5: Means ($\bar{x}$) and Standard deviations (σ_x) for each profile with respects to each sequence. A Friedman Anova ($\chi^2, df, p.value$) has been performed on each sequence.

4.4 Discussion

Taken together, our results suggest that the use of a washout filter improves the user's quality of experience. Counter-effects are removed, or at least reduced, which seems to make the haptic-audiovisual content more enjoyable. Nevertheless it appears that the tuning of the washout filter for maximizing the quality of experience depends on the user.

Group 1	T3	T2	T0
T2	0.0026	-	-
T0	$2.0e^{-07}$	$5.9e^{-07}$	-
T1	$2.0e^{-07}$	0.0015	$7.7e^{-08}$

Group 2	T3	T2	T0
T2	0.051	-	-
T0	$3.1e^{-07}$	$4.9e^{-07}$	-
T1	0.105	0.266	$8.9e^{-07}$

Group 2	T3	T2	T0
T2	$4.7e^{-05}$	-	-
T0	$1.3e^{-07}$	$1.0e^{-08}$	-
T1	0.23	$3.2e^{-06}$	$4.6e^{-07}$

Table 4.6: Pairwise comparison of each sequence for each profile using Wilcoxon test with Holm-Bonferroni correction.

All participants reported that one of the haptic feedback provided continuous movement (i.e. actuators were not moving between the effects which corresponds to the profile T1). This means that the counter-effects were actually not perceived. This feedback was the favorite for participants of Group 1 for which the higher the perceptual threshold used, the worst the ranking of the haptic feedback. The precise synchronization of the haptic feedback with the video seems to be a key component in the quality of experience for certain users.

On the contrary some participants have found the profile T1 less comfortable than the others. When the effect stopped, these users felt "being frozen" in a position different from the initial position, where all actuators are at rest. This perception may come from the muscles which do not support the same tension that in the initial position. Our model approximates the kinesthetic perception by the speed of joints and does not include the muscular perception. Then if the speed of the actuator is limited in a way to be not perceived, the off-center position is felt. This may explain why the profile T1 was not systematically preferred.

In line with this observation, participants of the Group 3 have classified the profile T2 as the best. They reported that they perceived counter-effects but it was not disturbing due to their weak intensity compared to the actual effects. The rendering was also more comfortable than with the profile T1. With the profile T3, the intensities of the counter-effects were too strong to be ignored and induced effects not coherent with the video.

From this experiment it appears that some users focus on the synchronization of the effects while others are more sensitive to the comfort of the position. But in both cases, counter-effects should be imperceptible. Not necessary from a kinesthetic point-of-view, but rather from a cognitive point-of-view. During a video viewing session, the user's attention is divided between visual, auditive and haptic stimuli. Haptic effects could be not perceived while above the perceptual thresholds. Hence, in addition to the user's sensibility, the perceptual thresholds for tuning a washout filter might be adjusted depending on the audiovisual context.

4.5 Application to a full video sequence

To complete this study we have tested the scalability of our approach in the context of a haptic movie. The new haptic rendering was applied on a movie enhanced with haptic effects. The movie used was "Sintel[1]" (see Figure 4.10), which total duration is around ten minutes. The haptic content was edited by a VFX artist thanks to a home made editor (see Chapter 5). Top part of Figure 4.11 shows an extract of the positions computed by the control model during a period of two minutes. Four counter-effects are present in this extract (represented by the grey rectangles).

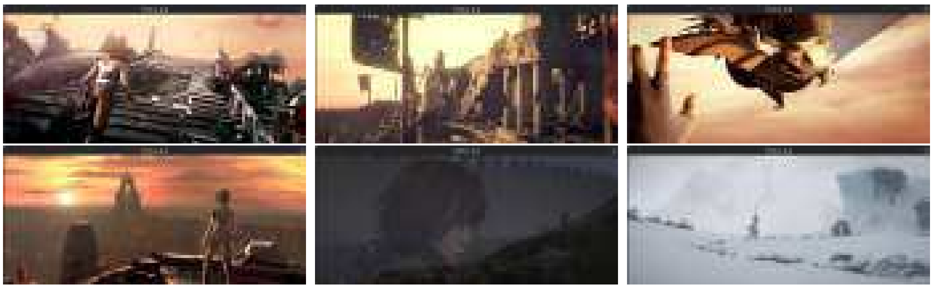


Figure 4.10: Screenshots of Sintel. Extract of the movie, from 4:44 to 7:02. Credit: Blender Fundation

The effects processed by our washout filter, with the T1 profile, are displayed on the bottom of Figure 4.11. The offset and scaling performed by the workspace optimizer are clearly visible. As expected the amplitude of the effects are reduced compared to the original content. The modification resulting from the perceptive optimizer is visible on the last effect, where the duration between two effects is the longest (420 to 425 seconds).

We informally asked to 11 volunteers to experience the two-minute extract of this haptic-movie. One time without washout filter, and another time with the washout and profile T1 (presented in a random order). They were asked to select their favorite sequence and to justify their choice. All of them reported that the haptic feedback is smoother with the washout filter, and this was appreciated. Almost two-thirds of the participants have clearly preferred the sequence with the washout filter enabled (7/11). The others found the amplitude of the haptic effects too small with the washout and preferred the other sequence. These first results follow those observed in the user study. The washout filter enables a smoother haptic rendering. A majority of users are focused on the synchronization of the haptic effects to the audiovisual content and prefer the use of a washout, whereas some others are more in demand for strong effects. But interestingly enough, this application proves that our haptic rendering works with real sequences.

[1]Credit: Blender Fundation. http://www.sintel.org

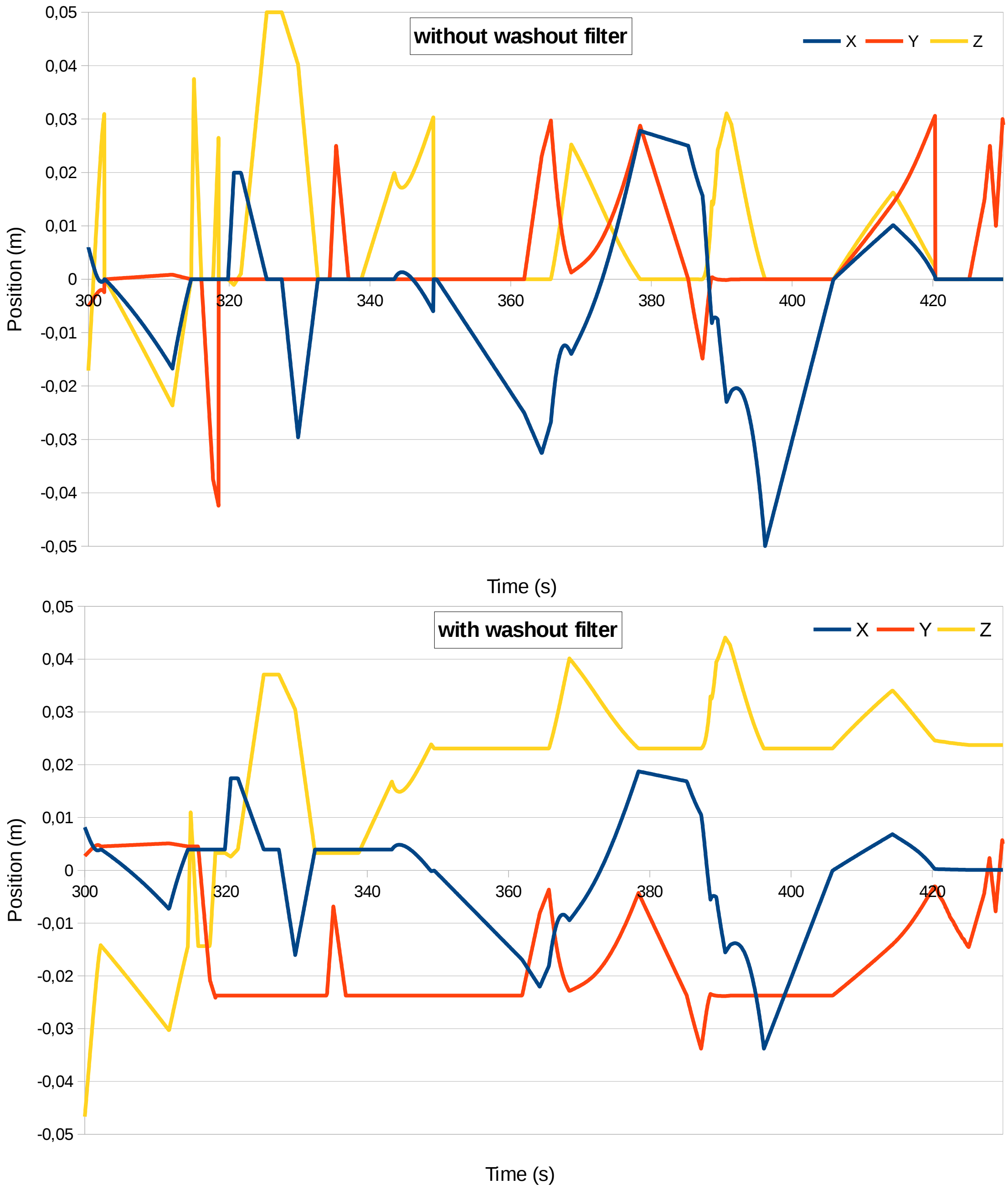


Figure 4.11: Extract of the haptic effects for the full video. Haptic effects without the washout filter are displayed on top, with the washout filter and the profile T1 on the bottom.

4.6 Chapter conclusion

In this chapter, we have presented a new haptic rendering algorithm for force-feedback device in a HAV context. It relies on a washout filter based on the kinesthetic perception. A user's body model is used to compute the angular speed of the joints of the moving limbs. These speeds are compared to perceptual thresholds to determine if the movement of the force-feedback device is felt. This movement can be then adjusted in order to make undesirable effects imperceptible. Moreover, three profiles with different perceptual thresholds have been designed to explore different implementation of rendering.

The results from a user study showed that the washout filter globally improves the QoE during video viewing. Besides, the results provided interesting insights regarding the tuning of such a washout filter. Parameters should be adjusted depending on the user's kinesthetic sensibility and perception of the audiovisual content. The haptic rendering was also applied to a full video sequence, showing the scalability of our approach.

Part II

Producing Haptic Effects: Tools and Techniques for Creating Haptic-Audiovisual Content

Chapter 5

H-Studio: an authoring tool for adding haptic and motion effects to audiovisual content

Contents

Despite the increasing interest for haptic effects in the context of audiovisual content, one main problem remains: how to design and when to insert haptic or motion effects in a given movie timeline so that the final effect may be relevant for the user? The selection of effects is strongly linked to the content (explosions, fast motions, camera effects) and some of them may be automatically proposed and placed on the timeline on the basis of the audiovisual content analysis. But the creative part of the author should provide a better final result, with optimal choice and placement of effects. It seems therefore necessary to provide content creators with authoring tools, easy-to-use and similar to their usual editing tools. But few editors already exists and they do not allow to intuitively edit 6DoF effects of motion.

This chapter describes such a tool and introduces a novel user interface dedicated to the creation of haptic and motion effects: the *H-Studio*. The authoring tool and its new features are detailed in section 5.1. In this work we focus on the creation of complex motion effects by proposing three editing methods. Two methods take advantage of a force-feedback device to enable the intuitive creation of motion effects. The third allows the import of real-world data. Such data are provided by a novel capture device we have designed, combining an inertial measurement unit and a video camera. Moreover, the authoring tool features the preview of motion effects rendered on a force-feedback device. A user study has been conducted to evaluate the quality of motion effect captured by our device, and more generally, the impact of the haptic feedback on the user's quality of experience. Protocol and results are presented in section 5.2. Finally conclusion is provided in section 5.3.

5.1 The authoring tool: H-Studio

We propose a new authoring tool to easily edit and preview haptic and motion effects. The interface of the editor, inspired from traditional video editing software, is composed of three main parts: a preview of the video, a timeline for the synchronization of effects and a menu with the parameters of the current effect (see Figure 5.2). So far, two types of effects are supported: vibration and motion. Two tracks are displayed above the timeline, each one is dedicated to one type of effect. The editor could easily be extended by the addition of more tracks associated to new haptic effects.

To create an effect, the user determines when it starts and stops on a track, then defines waypoints. A parameter menu allows to finely tune each waypoint and data are interpolated between them (linear interpolation). An effect can also be saved in a library in order to be reused.

Parameters for vibration effects are quite simple to edit: amplitude and frequency. The vibration is a sinusoidal signal created from these two parameters. The editor also proposes to automatically create vibrations from the audio track of the video content. The audio signal is directly used to represent the vibration effect. A filter is applied to adapt the signal to the capabilities of the vibrating device. This is a classical technique already used in the literature.

Motion effects are however less trivial to edit due to the 6DoF which all have to be set at a time: three linear accelerations (a_x, a_y, a_z) and three rotational speeds (w_x, w_y, w_z) [SACH10]. A motion effect can be formalized as:

$$M^t = [a_x, a_y, a_z, w_x, w_y, w_z]^t \tag{5.1}$$

We focus here on the edition of motion effects. We propose two methods relying on a force-feedback device to intuitively design motion effects (see Figure 5.1). Moreover, we also propose a new capture device to create a video augmented with motion effects. Captured data can directly be imported in the editor. Finally, whatever the method used, the motion effect can be previewed on the force-feedback device. The content creator can thus directly feel the synchronization and dynamics of the motion effects, without using a cumbersome motion platform.

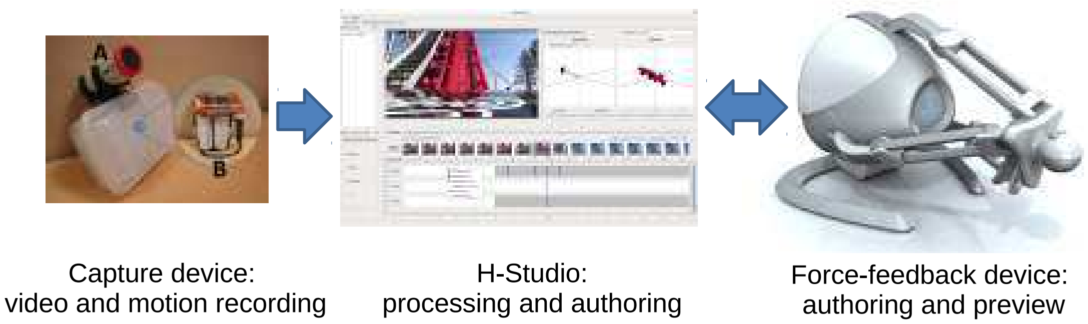


Figure 5.1: Overview of the *H-Studio*. Video and motion effects can be imported from our capture device, or manually edited thanks to a force-feedback device. This device can also be used to preview motion effects.

5.1.1 Manual motion effect edition

As it is not a simple task for the user to imagine a movement in 6DoF, we propose two methods to manually edit motion effects, taking advantage of a force-feedback device.

5.1.1.1 Waypoint edition using a force-feedback device

This first method enables a force-feedback device to edit motion parameters for each waypoint (see Figure 5.2). A waypoint is set at an instant t when the creator clicks on the timeline. An effect of motion M^t is associated to this waypoint. By manipulating the force-feedback device, the creator can directly set a direction vector and an orientation, which represent the acceleration (a_x, a_y, a_z) and rotational speed (w_x, w_y, w_z) for the instant t. For example, if the creator moves the actuator forward, the motion effect is set as a forward acceleration. The bigger the amplitude of the creator's movement is, the bigger the amplitude of the motion effect is.

In our implementation we rely on a Novint Falcon, a low-cost 3DoF force-feedback device [NOV]. Only the direction or the orientation can be edited at a time. This approach should easily be extended to a 6DoF device.

5.1.1.2 Trajectory recording from a force-feedback device

Creating a complex effect might be tedious with the previous technique as motion parameters must be edited for each waypoint. The second proposed method is to record a trajectory thanks to the force-feedback device without using waypoints. The creator directly moves the device while the video is playing for a duration previously defined. The drawn trajectory represents the motion effect M. This solution is more intuitive although the synchronization with the video could be less easy.

When the user starts the recording, the positions of the actuator are sampled at 30Hz and stored in a simple "csv" file. A timestamp is associated to each sample. Here again we rely on a Novint Falcon which allows to record 3DoF motion effects. But this technique is easily extendable to a 6DoF force-feedback device.

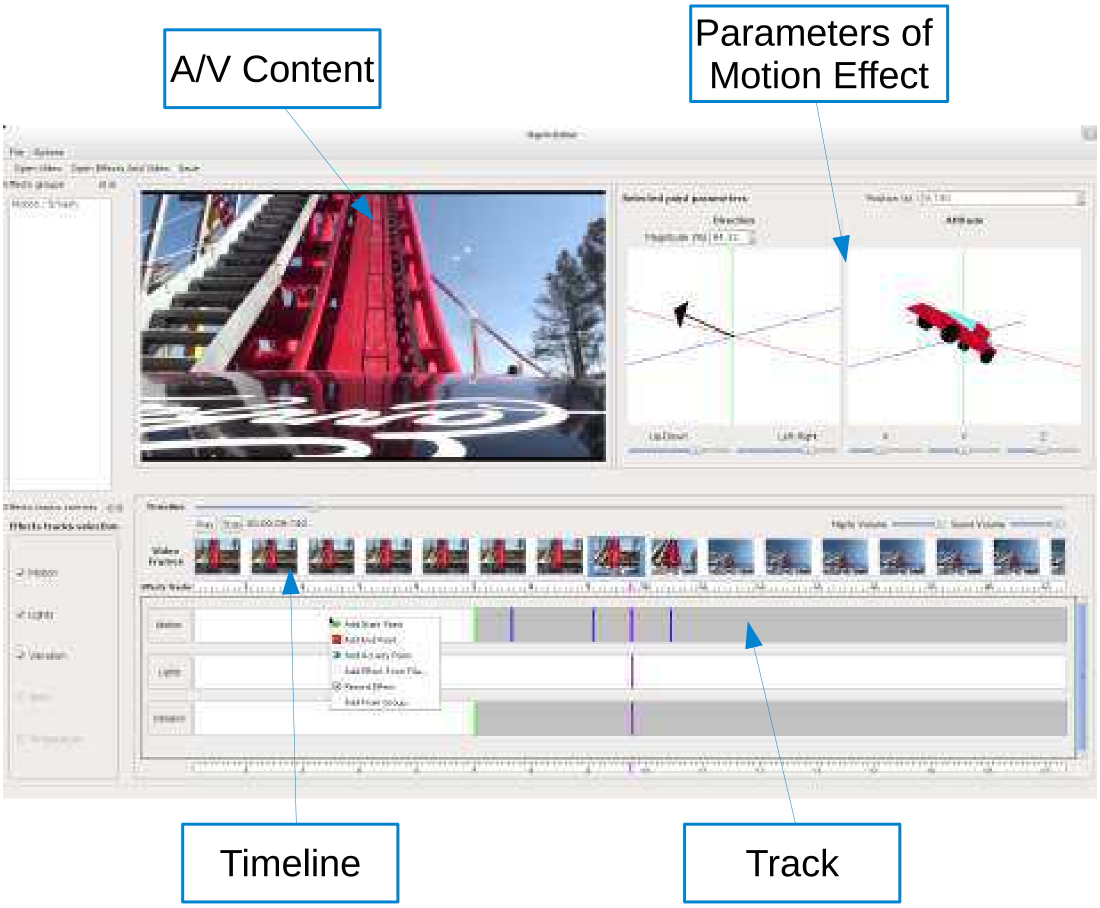


Figure 5.2: Screenshot of *H-Studio*. A motion effect is being edited: direction (represented by an arrow) and orientation (represented by a car) are defined at an instant t.

5.1.2 Automatic creation of motion effect

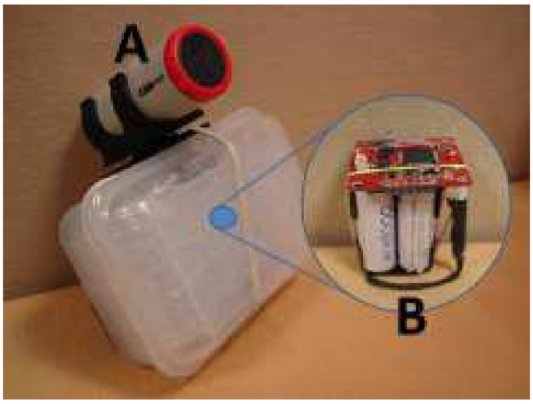
The previous method allows to create complex trajectories, but designing a highly realistic motion effect as it would be felt in the real world is not straightforward. Then we have developed a new capture device to record motion effects during the shooting of the video.

5.1.2.1 Capture device

The capture device is a combined system making use of an inertial measurement unit (IMU) and of a high definition camera dedicated to sportive activities. A complete integrated prototype combining the IMU, its battery and the camera has been developed. As the system is designed to be fixed on an actor (first-person point of view recording), it is robust enough to resist to different conditions of recording (see Figure 5.3b).

The IMU we chose is the Ultimate IMU board which combines an ADXL345 accelerometer, an ITG-3200 gyroscope and a HMC-5843 electronic compass (see Figure 5.3a). The first component records the 3-axis accelerations of the board $\mathbf{a_r}$, the second one quantifies the rotational speed of the board around its 3 axes, $\mathbf{w_r}$, and the last component allows a geocentric orientation by giving an estimation of the local magnetic field, $\mathbf{c_r}$. An additional micro-SD memory card may be embedded on the board and allows the recording of the three raw signals. A dedicated middleware has been developed and uploaded onto the Ultimate IMU to set the recording process of $\mathbf{a_r}, \mathbf{w_r}$ and $\mathbf{c_r}$ to 30Hz. A timestamp is associated to each sample. A filtering is necessary to reduce the noise of the original signal. For practical reasons, the filtering is actually performed on the IMU. More precisely, the three sensors data were natively sampled at 200Hz but due to limitations with the writing speed on the embedded micro-SD, samples were averaged and down-sampled at 30Hz. This averaging step results in a low-pass filtering of the raw signal.

Complementary, a Camsports HDS-720p was selected to record the scene corresponding to the current point of view of the actor (see Figure 5.3a). The camera is a HD bullet camera. It uses a 120 degrees wide-angled lens and integrates a built-in 4GB memory chipset. The spatial resolution of this device is 1280×720p at a frequency of 30fps. It is water-proof and is able to handle harsh environments. It finally integrates a mono-channel microphone.



(a) Prototype composed by (A) a Camsports HDS-720p and (B) an Ultimate IMU board.

(b) The prototype is fixed on an actor's chest and records motion on three axes.

Figure 5.3: Overview of the device capturing both video and motion.

5.1.2.2 Processing of the captured motion signals

Both the video and the motion effect captured by the device can be imported into the *H-Studio*. To synchronize the IMU and the camera, which do not offer possibilities of external synchronization, a mechanical trick is used (very similar to the audiovisual synchronization techniques traditionally used in movie making). Before each record, three little pats are given on the prototype which cause a fast and big peak in both the acceleration signals of the IMU and the audio stream of the camera. Basic signal processing techniques are then used to make those peaks match in both signals (variance-based threshold).

The signal recorded by the IMU has to be processed in order to be rendered on a haptic device. The main processing to apply is linked to the gravitational component $\mathbf{g}$ included in the raw acceleration $\mathbf{a_r}$. This latter is quite important regarding the other external sources of acceleration and can mask some useful information needed to render a motion feeling. The board orientation is estimated using the approach described by Sabatini [Sab06]. This latter especially combines the use of a quaternion-based representation of the board attitude and a dedicated extended Kalman filter to estimate the board orientation by merging the information coming from the three sensors (gyroscope, accelerometer and magnetometer). This operation allows to estimate the direction of the gravity (vertical) $\mathbf{n}_g[k]$ in the accelerometer frame (frame $\mathbf{A}$) at each time sample k. The raw acceleration vector a_r is therefore updated by removing the quantity, $\|\mathbf{g}\|\mathbf{n}_g[k]$, from each sample $\mathrm{a_r}[k]$.

The new acceleration vector $\mathbf{a}[k] = \{a_x[k], a_y[k], a_z[k]\}^t$ may be formalized, at each time sample k, by:

$$\mathbf{a}[k] = \mathbf{a_r}[k] - \|\mathbf{g}\|\mathbf{n}_g[k] \tag{5.2}$$

In our context the captured motion effect is thus defined by:

$$M^t = [a_x, a_y, a_z, w_{rx}, w_{ry}, w_{rz}]^t \tag{5.3}$$

Extra operations may be performed for enhancing the signal for a better rendering for the end-user. They may be simple operations to remove artifacts or artificial modulation (reduction or amplification) of some parts of interest in the M^t to underline specific haptic events.

5.1.3 Preview of motion effects

Once edited, the creator may want to preview the motion effect. However, end-devices (such as motion platforms) are not always available or would not be convenient for a quick preview. In line with the approach proposed by Ouarti et al. [OLB09], we propose to use a force-feedback device to preview these effects. Figure 5.4 shows the setup used to render the motion effects.

An open-loop rendering system was introduced to display motion effect M^t. Our implementation relies on the Novint Falcon which proposes 3DoF force-feedback rendering. Only the acceleration composing the motion effect M^t was rendered as a force vector F, defined as $F[k] = \{F_x[k], F_y[k], F_z[k]\}^t$.

To be rendered on the haptic device, an axis permutation of the signal M has to be performed to align the axes of the accelerometer (frame $\mathbf{A}$) with the axis of the device (frame $\mathbf{D}$). The associated permutation matrix is termed $\mathcal{P}_{\mathbf{D}}^{\mathbf{A}}$. Besides a scaling of the data is necessary to adapt the amplitude of the signal M to the workspace of the device. The force rendered by the haptic device may be finally formalized by:

$$F = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{bmatrix} \mathcal{P}_{\mathbf{D}}^{\mathbf{A}}(M) \tag{5.4}$$

In our context, the matrix simply switches the axes Y and Z of **A** in **D**. A complementary step reverses the Z-axis as the force-feedback device is placed in front of the user and it is supposed to pull the user's hand when the recorded acceleration is positive on the z-axis. The scaling factors s_x, s_y and s_z for each axis are assumed to be constant (independent of the time sample) and empirically set according to experimental feedback.

During the haptic rendering, the force F is computed for each sample M and oversampled (piecewise constant interpolation) to meet the requirements of the 1kHz haptic rendering loop frequency.

5.2 User study

The user study is focused on the evaluation of captured motion effects when previewed on a force-feedback device. We wanted to evaluate the realism of the motion effects captured by our device which is a key factor in the design of such effect. Also we evaluated the preview of such effects on one force-feedback device. More generally, we wanted to study the impact of the haptic feedback on the user's quality of experience (QoE [Jai04, Kil08]). In our context, the QoE may be defined as the measure of the user' subjective experience with haptic-audiovisual content. The realism is then a component of this QoE. Our hypothesis is that the haptic feedback improves this experience, even with a limited setup.

15 participants have taken part to the experiment. They were aged from 21 to 59 ($\bar{x}$=27.8 σ_x=9.7), nine were Male, one participant was left-handed, eight never used a force-feedback device. The whole experiment lasted from 30 to 40 minutes. Each participant was first introduced to the Novint Falcon and given a demonstration of its force capabilities. This step aimed to reduce the "surprise effect" for novice users. Participants were asked to passively experience each stimulus (see Figure 5.4 and Section 5.2.1) and then answer a questionnaire (see Section 5.2.3). A post-test questionnaire with open questions was also submitted in order to collect more details about the users' feelings.

5.2.1 Capturing test sequences

Our motion capture prototype was used to create several samples of audiovisual content enriched with motion effects. We identified four scenarios to represent different kinds of motion feelings (Figure 5.5). The prototype was placed on an actor's chest and we obtained the following contents:

Figure 5.4: A participant experiences one of the video sequence enriched with motion effects.

1. **Bike**. The objective of this scenario is to capture low-amplitude movements. The actor is performing outdoor cycling and a succession of vertical movements with small amplitude is captured. (duration 61s).

2. **Horse**. In this case the actor is riding a galloping horse and feels recurrent top-down movements. High-amplitude vertical movements are captured. (duration 60s).

3. **Car turning**. In this scenario, the actor is inside a car engaged in a roundabout. The centrifugal force makes him feel pushed on a side. The captured motion is felt as strong and long. (duration 45s).

4. **Car Braking**. This last scenario aims to capture a strong punctual movement. The actor is in a car strongly braking and feels a strong force pushing him forward during few seconds. (duration 75s).

(a) Outdoor cycling (b) Horse riding (c) Car turning (d) Car braking

Figure 5.5: Tests Scenarios.

5.2.2 Variables

In order to evaluate the user's QoE for each sequence we defined three types of haptic feedback to be rendered with the video:

1. **Realistic Feedback**. The captured haptic feedback, consistent with the sequences.

2. **No Feedback**. Only the audiovisual content is displayed. The goal of this condition is to measure the QoE of a classical audiovisual content. This will be used as a reference to evaluate the interest of a haptic feedback for a video.

3. **Random Feedback**. A random haptic feedback made of a low-pass filtered white noise (cutoff frequency $F_c = 0.5Hz$) with the same length and amplitude than the consistent haptic feedback. This feedback is not consistent with the video and will be used to evaluate the interest of providing a realistic haptic feedback.

Combining the whole set of possibilities, **12 conditions** (4 videos sequences $\times$ 3 types of haptic feedback) are obtained and were tested in each experiment in order to evaluate the QoE, our independent variable. These conditions were presented in a random order to the participants.

5.2.3 Measures

A questionnaire was designed to evaluate the QoE of a video enriched with haptic feedback. It was built around the **Presence** [WS98] and **Usability** [TA08] concepts (similarly to the questionnaire used in Chapter 3).

Presence aims at measuring how much the user feels being physically situated in a virtual environment. Witmer and Singer [WS98] identified four factors to determine the presence: Control, Sensory, Realism and Distraction. "Control" determines how much the user can control and modify objects within the virtual environment. "Sensory" characterizes how each sensory modality is solicited during the interaction. "Realism" describes how much the environment is realistic and consistent with user's representation of the real world. "Distraction" identifies how much the user is disturbed by the apparatus used to create the virtual world. From this definition we focused on two factors: *Realism* and *Sensory*. As the user is passive with our system, Control factor was not relevant here. Moreover we did not measure Distraction in our QoE questionnaire, but this aspect was interesting and was evaluated in the post-test questionnaire.

Usability is defined by the norm ISO 9241-11 and aims at measuring how easy a system is to use. Three factors composed this concept: Efficiency, Effectiveness and Satisfaction. This latter measures how well the user enjoyed the system. "Effectiveness" means how well a user can perform a task while "Efficiency" indicates how much efforts are required. These two factors were not totally suitable for our system in the sense that it was not designed to perform a task. We preferred to use the term of *Comfort* to measure how well was the system to provide feedback. *Satisfaction* was however fully relevant in our situation.

Hence, the QoE of our system was evaluated by four items : Realism, Sensory, Comfort and Satisfaction (see Table 5.1). We defined only one question by item supposed to be rated on a five-point Likert-scale. The QoE is computed by the sum of these four items. This way the QoE questionnaire is easy to fill in and can be submitted for each condition.

Factor	**Question**
Realism	How much did your experiences in the virtual environment seem consistent with your real-world experiences?
Sensory	How much did the haptic feedback improve the interaction?
Comfort	Was the system comfortable?
Satisfaction	Was the system pleasant to use?

Table 5.1: QoE Questionnaire. Each question is rated on a 5-point Likert-scale from 1 (Not at all) to 5 (Totally).

5.2.4 Results

The collected data were four notes (associated to Realism, Sensory, Comfort and Satisfaction; from 1 to 5) for each condition per participant. The sum of these notes gives the score for the QoE per conditions per participant. The normality of the distributions was tested with the Shapiro-Wilk test and was rejected most of the time. Hence non-parametric tests were used to analyze the results presented in this section (Friedman Anova and Wilcoxon test with Holm-Bonferroni correction).

We first looked at the result for all the video sequences combined (see Table 5.2). $QoE_{Realistic}$ ($\bar{x} = 15.3$, $\sigma_x = 2.6$) has the highest score, followed by QoE_{Random} ($\bar{x} = 10.2$, $\sigma_x = 1.6$) and QoE_{No} ($\bar{x} = 7.5$, $\sigma_x = 2.1$). This result, depicted on Figure 5.6, is significant according to the Friedman Anova ($p < 0.05$). We have also observed that the QoE for each individual sequence follows the same pattern (Figure 5.7).

Figure 5.6 shows the mean score for each item of the QoE, for the three feedback conditions. The more realistic the feedback, the higher the Realism, Sensory and Satisfaction scores (see Table 5.2). These results are also significant according to Friedman Anova and Wilcoxon tests (see Table 5.3). However Comfort appears to be relatively stable all along the experiment ($\bar{x}_{None} = 2.9$, $\bar{x}_{Random} = 3.2$, $\bar{x}_{Realistic} = 3.6$). According to Wilcoxon test they are indeed statistically equivalent.

Finally we observed that the $QoE_{Realistic}$ remains the same for those who never used a Novint Falcon ($\bar{x} = 15.5$ $\sigma_x = 2.9$) and for those who did ($\bar{x} = 15.25$ $\sigma_x = 2.5$). The expertise of the participant do not affect the result significantly (Wilcoxon test, $p = 0.77$).

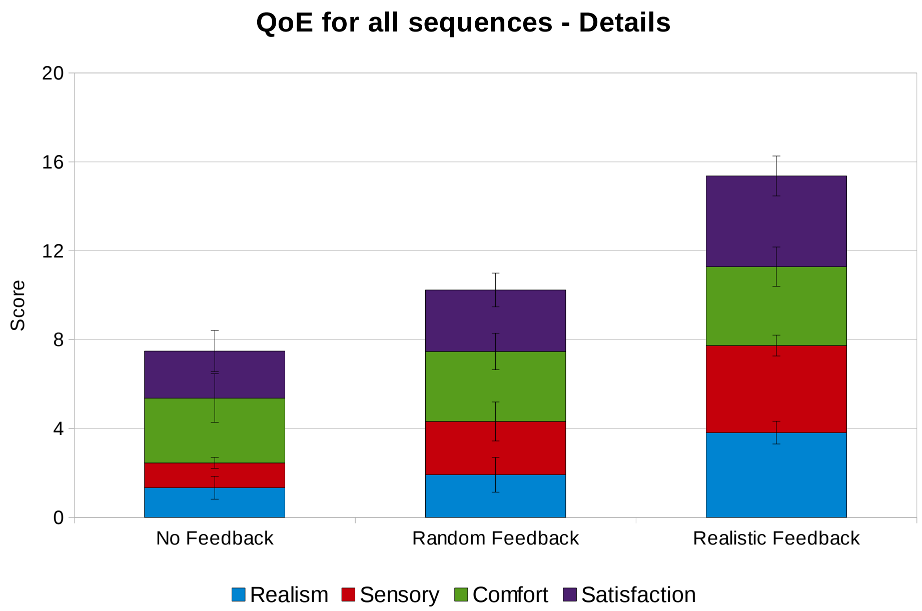


Figure 5.6: QoE of each haptic feedback and details of the components. The Comfort component of the QoE remains the same whatever the feedback perceived. However the three others increase with no feedback, random feedback and realistic feedback respectively.

Factor / **Model**	QoE	Realism	Sensory	Comfort	Satisfaction	
None	7.4833 1.5597	1.3333 0.5147	1.1167 0.2476	2.9167 1.0965	2.1167 0.9252	$\bar{x}$ σ_x
Random	10.2333 2.1391	1.9167 0.5147	2.4 0.4706	3.1500 0.8854	2.7667 0.8987	$\bar{x}$ σ_x
Real	15.3667 2.6184	3.8167 0.7819	3.9167 0.8746	3.5500 0.8194	4.0833 0.7599	$\bar{x}$ σ_x
F. Anova	24.7119 2 $4.30e^{-6}$ ***	24.0339 2 $6.04e^{-6}$ ***	26.678 2 $1.61e^{-6}$ ***	8.7917 2 0.0123 *	23.0526 2 $9.867e^{-6}$ ***	χ^2 df p sig.

Table 5.2: Means ($\bar{x}$) and Standard deviations (σ_x) for each model with respects to each factor. A Friedman Anova ($\chi^2, df, p.value$) has been performed on each factor.

5.2.5 Discussion

The main result of this study is that the motion effect captured by our device is perceived as realistic and such effect improves the QoE. Moreover the expertise of participants with a force-feedback device does not affect the QoE. This observation let us think that our main result is not due to a “surprise effect” and that the setup is suitable for nonexpert users.

This study has also brought interesting results regarding the design of motion effects

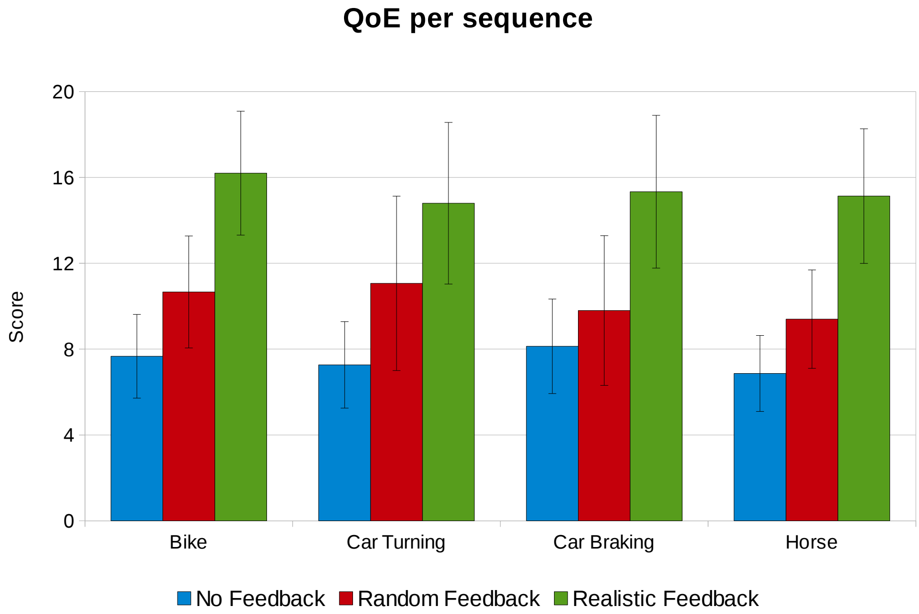


Figure 5.7: QoE of each sequence and haptic feedback. For each sequences, participants found that a realistic haptic feedback improves the experience. Interestingly a random feedback was more appreciated that no feedback.

QoE	No	Random
Random	0.0016	-
Real	$1.2e^{-5}$	0.0002

Realism	No	Random
Random	0.004	-
Real	$1.2e^{-5}$	$2.9e^{-5}$

Sensory	No	Random
Random	$7.3e^{-6}$	-
Real	$5.0e^{-6}$	0.0001

Comfort	No	Random
Random	0.48	-
Real	0.12	0.48

Satisfaction	No	Random
Random	0.06	-
Real	0.0002	0.002

Table 5.3: Pairwise comparison of each model for each factor using Wilcoxon test with Holm-Bonferroni correction.

for video viewing context. QoE increases with haptic feedback and more particularly with haptic feedback consistent with audiovisual content. However the low score obtained by sequences without haptic feedback (No Feedback condition) can be in part due to our experimental protocol. Whatever the condition, participants were asked to hold the force-feedback device in their dominant hand. Thus they might have been frustrated by the absence of feedback. Obviously if there is a haptic device, people are expecting haptic feedback.

We have also observed that haptic feedback may change user's perception of the audiovisual content, especially if the meaning of the video is ambiguous. For instance one cannot see a bike in the bike sequence although a head of a horse is visible in the horse sequence as well as a part of a car in the two car sequences. During the experiment a participant thought that the bike sequence represented a buggy riding video because he felt that the haptic feedback (Realistic Feedback condition) was closed to his own buggy driving experience. Thus it appears that users build a mental representation of the multimedia content consistent with their own experience, and it is interesting to see how haptic feedback can influence this representation when audiovisual content is ambiguous.

Another interesting behavior was observed while participant experienced video enriched with random feedback. Most of them tried to find a meaning for this haptic feedback, from their own personal experience. This observation may explain higher QoE for random feedback than for no feedback. The phenomenon was particularly highlighted in the Car Turning and Car Braking conditions. Several participants supposed that the haptic feedback was mapped to the gear shift of the car. This can also explain why QoE for Random Feedback in these two conditions is better than in Bike and Horse conditions.

Finally participants reported in the post-test questionnaire to feel comfortable all along the experiment although the position of the arm and the hand-grip were reported as quite uncomfortable. This setup is obviously not suitable for watching a two-hour movie but is suitable in a previewing context.

This first user study yielded interesting results for designing motion effects. Therefore, research efforts are necessary to determine when the user perceives a haptic feedback as consistent or not with an audiovisual content. This will help to finely design effects necessary to trigger an immersion feeling. The evaluation of the two other editing methods is also necessary. A usability study should be conducted to evaluate the strengths and weaknesses of each method.

5.3 Chapter conclusion

This chapter introduces the *H-Studio*, a new authoring tool to create haptic and motion effects for audiovisual content. Three methods are proposed for the edition of motion effects. The first method enables the manual edition of motion parameters (acceleration and rotation speed) thanks to a force-feedback device. The second method allows to directly draw a trajectory using this device. The trajectory represents the motion effect. The third method consists in the import of real motion captures. Such captures can be performed by a new device we proposed, allowing to record both audiovisual content and motion effects. The authored effects may be easily previewed, which enables an iterative design process. Our playback system relies on a force-feedback device to make the user feel the motion effects while watching the video.

Finally we have conducted a user study and presented a questionnaire to evaluate users' quality of experience when previewing captured effects. Results show that the user experience increases with a realistic haptic feedback. Besides they bring useful insights for designing motion effects.

This new tool could simplify the creation of haptic-audiovisual content. This brings a new way to experience multimedia content and can enhance many viewing contexts such as movies, extreme sports videos or video games.

Chapter 6

Toward haptic cinematography: enhancing movie experience with haptic effects based on cinematographic camera motions

Contents

Today, haptic-audiovisual (HAV) where users see, hear and physically feel the content, is mostly experienced in "4D cinemas" or amusement parks. But new devices are developed to bring this technology to consumers. A typical example is the seat developed by the D-Box company. With the provision of new haptic devices, appears the necessity to create new HAV contents, and to design new modalities for the creation of haptic effects. Haptic effects often represent physical events occurring in an audiovisual scene. However many other aspects could be enhanced.

In this chapter, we propose to consider haptics as a new component of the film-maker's toolkit. We dubbed this approach *Haptic Cinematography*. A taxonomy of haptic effects that classifies potential haptic effects for audiovisual content and the context in which they may be used is first presented in section 6.1. Among the possible effects, the coupling of haptic effects with cinematographic camera motions has not been addressed. Hence we introduce a new type of haptic effect related to camera motions (referred as camera effects) that are used by movie makers to convey meaning or to create emotion. We propose two models to render camera effects on haptic devices. The first model is designed to make the viewer feel the movement of the camera, the second provides a haptic metaphor related to the semantics of the camera effect. The proof-of-concept is described in section 6.2, followed by the user study in section 6.3. Discussion and conclusions are provided in sections 6.4 and 6.5.

6.1 Haptic cinematography

Cinematography encapsulates both the art of making movies and the associated techniques (camera work, staging, lighting, sound, montage, etc.) [TB09]. In order to improve users' experience, many others effects have been added: special visual effects, spatialized sound, 3D technology, etc and we believe that haptics should also be included in the filmmaker's toolkit.

We introduce the concept of *Haptic Cinematography* which represents the techniques to create haptic effects in order to produce a HAV content and organize effects in a taxonomy (see Figure 6.1).

6.1.1 Taxonomy of haptic effects

A parallel can be drawn between the role of haptic effects and the one of audio in movies: audio is used for increasing the realism (sound effects) but also to create ambiance (music). These two categories of audio content are known as diegetic sounds, a sound for which the source belongs to the diegesis (the recounted story), and non-diegetic sounds, a sound for which the source is neither visible nor implied in the action, typically such as a narrator's comment or mood music [TB09]. In a similar way, haptic effects can be classified into diegetic and non-diegetic effects.

Diegetic haptic effects can enhance physical events happening (and usually visible) in the audiovisual content in a similar way to how haptic effects are used in virtual reality applications. Two subcategories may be identified: local or global. Local effects are associated to one object in the scene: e.g. force-feedback [OO03] or vibrations [KCRO10] related to events occuring with an onscreen character or vibrations representing the position of the ball in the soccer game [uR08]. Global effects refer to effects related to the environment. This could be vibrations associated to an earthquake in a

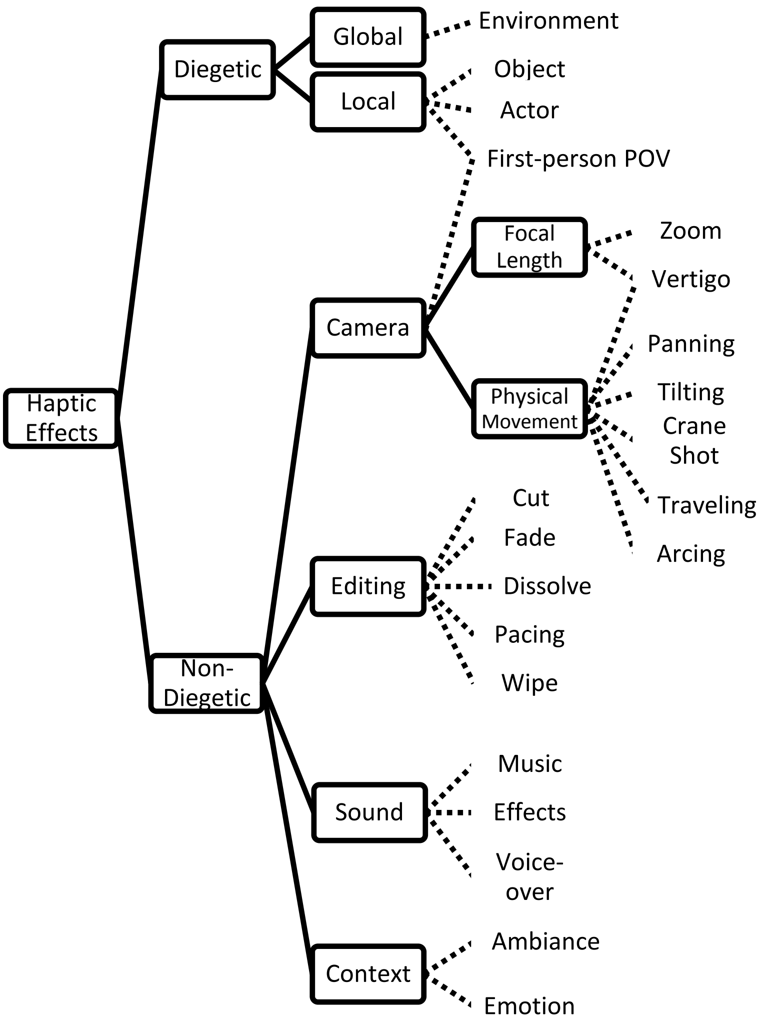


Figure 6.1: Taxonomy of haptic effects for audiovisual content. Items in boxes are categories and those linked with dash lines are examples.

movie or a system allowing users to touch the objects within the scene (see Cha et al.'s touchable TV [CES09]).

Non-diegetic effects refer to elements not attached to the fictional world depicted by the story. Davenport et al.'s have proposed a model of the shot which includes non-diegetic elements [DSP91]. From this model, we identified four categories of non-diegetic haptic effects. The first category of effects is related to non-diegetic sounds (i.e. music, voice-over, etc.). Here haptic effects would highlight particular sound effects or music [LC13]. In a second category, haptic effects underline the context, i.e. the ambiance or emotion (Lemmens et al.'s jacket [LCB+09]). More generally the design of such effects would take advantage of research results in affective haptics to convey emotion through haptic feedback [TNP+09]. A third category contains effects related to the camera parameters, focal length and physical movement, which are used by movie makers to achieve visual effects. Editing techniques could be used in a similar way. The editing process is another tool employed by movie makers to convey emotion or meaning [TB09]. For example the "pacing", the rhythm due to the succession of shots, may create tension. A haptic effect could follow this rhythm to increase the tension.

To the best of our knowledge, no work relied on the camera or editing to create haptic effects. Similar techniques may exist in the field of virtual reality where the user can manipulate the camera. But our proposal fundamentally targets a different context: the association of haptics to cinematographic elements. There is no interaction and the aim is more to increase the cinematic experience than only moving the user's point of view. These cinematographic techniques are intensively used to convey meaning or emotion. Our hypothesis is that haptic feedback may underline these effects and therefore improve the quality of the video viewing experience. To illustrate this approach we focus on enhancing camera effects with haptic effects.

6.1.2 Camera effects

A camera effect consists in modifying the camera parameters such as the position of the camera or the focal length to obtain a specific visual effect [TB09]. If there is no strict rule, camera effects are generally associated to a specific purpose. For example, the "Vertigo" effect, also known as "Dolly Zoom", has been democratized by Alfred Hitchcock in his Vertigo movie released in 1958. This effect is a combination of a zoom-out and a forward movement of the camera. The result is that the environment around the framed object is being distorted, which induces a sensation of vertigo.

We identified seven main representative camera effects from the cinematography literature [Mas98, TB09]: three movements (Crane Shot, Arcing and Traveling), two rotations (Dutch Angle and Tilting), one modification of the field of view (Zoom) and Vertigo. Table 6.1 describes how they are created and the purpose for which they are commonly employed.

6.1.3 Haptic effects based on camera effects

We designed haptic effects to underline the visual effects achieved by the camera motions: the vertigo sensation of the Vertigo effect, the feeling of instability triggered by a Dutch Angle or the movement of the camera during a Traveling.

We proposed two different models to render haptic effects based on camera effects. The first one aims at making the user feel the movement of the camera (a zoom is considered as a forward movement). This model is called *Cinematic Model*. We assume that information about the position, pose and field of view of the camera is available and can be used to drive a haptic device. The second model renders a haptic effect which is related to the purpose of the cinematographic effect (see Table 6.2). We dubbed this model *Semantic Model*. In this case the effect is manually authored and would be designed as a metaphor for the cinematographic effect.

Camera Effect	Description	Purpose	Camera Parameter
Crane Shot	Vertical movement such as a lift-off	Feeling of omniscience over the characters	y_c, ϕ_c
Dutch Angle	Tilting to a side	Underline physiological uneasiness or tension	ψ_c
Arcing	Circle movement around the framed object	Increase the importance of the scene	x_c, z_c, θ_c
Traveling	Lateral movement	Follow an object or actor	x_c
Tilting	Rotation in a vertical plane from a fixed position	End with low angle: feeling of inferiority regarding the framed object	ϕ_c
Zoom-in	Modification of the focal length	Attract attention toward an object	γ_c
Vertigo	Zoom-out while the camera moves forward	Sensation of vertigo or strangeness	z_c, γ_c

Table 6.1: Cinematographic camera effects. They are typical movements along one or more degrees of freedom and/or a modification of the focal length and they are usually associated to a specific meaning [Mas98, TB09]. The last column indicates which parameters of Equation 6.1 are modified in order to generate the effect.

Both models convert the camera effect into a haptic feedback. Then their implementation depends on the targeted haptic device. But the concept is applicable to any type of haptic device: force-feedback devices, tactile devices or even motion platforms.

6.2 Proof-of-concept

To evaluate the relevance of our approach, we have created seven video sequences illustrating the camera effects listed in Table 6.1. Then our two models were implemented and designed to render effects on the *HapSeat*, a novel haptic device which simulates sense of motion (see Chapter 3).

6.2.1 Audiovisual content

As already mentioned in the related work section, there are several ways to generate a video augmented with motion data: camera properties may be captured during production, they may be extracted from metadata in the AV content or they may be computed from image processing algorithms [Tho06].

Here a 3D engine has been used to generate video sequences illustrating the seven camera effects. We used a classical camera model to represent the position of the camera in space (Cartesian coordinates x_c, y_c, z_c), its orientation (three Euler angles ϕ_c, θ_c, ψ_c) and the value of its field of view, γ_c, for each instant t [CON08]:

$$C^t = [x_c, y_c, z_c, \phi_c, \theta_c, \psi_c, \gamma_c]^t \tag{6.1}$$

The 3D scene shows two characters animated with an idle behavior in a building (see Figure 6.2). The scene is voluntarily neutral to highlight the camera effect and to avoid potential distracting elements. The cinematographic effects were produced by modifying the camera parameters. For example a Traveling is a modification of the x_c parameter or a Tilting is a change of the ϕ_c parameter (see Table 6.1). The duration of a sequence was seven seconds: the camera stayed still for the first second, then camera parameters were modified in a way to produce a continuous effect during five seconds and finally it stayed still again for one second (hence reproducing the classical usage of cinematographic camera motions in movies). The screenshots of the created sequences are available in appendix C.

Figure 6.2: Screenshots of the Crane Shot sequence. The viewpoint displayed at the beginning of the sequence is modified by the movement of the camera (from left to right pictures).

6.2.2 Haptic device: the HapSeat

Haptic effects were rendered on the *HapSeat* (see Chapter 3). In the remainder of this chapter the following notation is used. The actuators near the head, left hand and right hand are labeled H, LA, and RA. Their central positions in their workspaces are named respectively G_H, G_{LA} and G_{RA}, G being the center of the space. The size of the workspace of one actuator is 10×10×10 cm.

6.2.3 Cinematic model

The purpose of this model is to mimic the movement of the camera for which all parameters are available. It is an extension of the *Geometrical* model described in Chapter 3. The command law to control one local actuator A is formulated in terms of displacement from its initial and central position G_A to the new position G'_A:

$$\overrightarrow{G_A G'_A} = f(\vec{T}, \vec{R}, \vec{F}) \tag{6.2}$$

where

$$f(\vec{T}, \vec{R}, \vec{F}) = \frac{\|\vec{T}\|\vec{T} + \|\vec{R}\|\vec{R} + \|\vec{F}\|\vec{F}}{\|\vec{T}\| + \|\vec{R}\| + \|\vec{F}\|} \tag{6.3}$$

and

$$\vec{T} = \begin{bmatrix} k_x & 0 & 0 \\ 0 & k_y & 0 \\ 0 & 0 & k_z \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \tag{6.4}$$

$$\vec{R} = (R_x(m_x\phi_c(t))R_y(m_y\theta_c(t))R_z(m_z\psi_c(t)) - I_3)\overrightarrow{GG'_A}) \tag{6.5}$$

$$\vec{F} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & s_z \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \gamma_c \end{bmatrix} \tag{6.6}$$

The function f is the combination of three vectors $\vec{T}$, $\vec{R}$ and $\vec{F}$ which respectively uses the positions, pose and focal length parameters of the camera model (Equation 6.1). k_x, k_y, k_z, m_x, m_y, m_z, s_z are some scaling factors to map the motion of the camera in the workspace of the actuator. R_x, R_y and R_z are the 3D rotation matrices around their respective X, Y and Z axes and I_3 is the identity matrix of R^3.

From this equation, the new application points G'_H, G'_{LA} and G'_{RA} are computed from the initial points G_H, G_{LA} and G_{RA}. The scaling factors are computed to use the workspace of each actuator in an optimal way, by finding a compromise to avoid any saturation while using the largest space available. The computation of those scaling factors is performed by a preprocessing step consisting in finding the maximal amplitude of displacement rendered by the three different actuators.

The output of this model is specific in the case of the Vertigo effect. The effect is composed by a combination of a forward movement (input of Equation 6.4) plus a zoom-out (which is considered as a backward movement by Equation 6.6). Thus the model produces no movement for this effect. For the other cases the user will follow the movement of the camera described in Table 6.1: for the Zoom-in, the user feels a forward movement (see Figure 6.3); for the Dutch Angle, the user feels a rotation (left actuator goes down while the right one goes up); for the Traveling, the user feels a lateral movement; etc. The output for all the sequences is provided in appendix B.

6.2.4 Semantic model

The second model aims at evoking the purpose of the camera effect. For example, the Dutch Angle is often used to show that something strange is happening (Table 6.1). The associated haptic effect should therefore highlight this sensation of strangeness.

Camera Effect	**Metaphor**	**Description**	**Implementation**
Crane Shot	Flying away	User feels several up and down movements as a bird taking off.	Actuators are going up then down with an increasing intensity.
Dutch Angle	Instability	User sways from left to right, as on a boat.	Left actuator goes up while the right one goes down and vice versa.
Arcing	Intensification	User's hands are getting closer in a movement to represent a concentration.	All actuators are moving towards the center G.
Traveling	Crab walk	Hands movement mimic a crab walk following the camera movement.	Right actuator move toward the right. Then it slightly goes back to its initial position while the left actuator move toward the right. And so on.
Tilting	Inferiority	User's hands and head go down to make the user feel smaller than the framed object.	All actuators go down.
Zoom-in	Walk forward	User's hands movement mimic a forward walk.	Similar to crab walk except that the actuators move forward.
Vertigo	Vertigo	User's hands move away from each other as if the environment is being extended.	All actuators are moving away from the center G.

Table 6.2: Semantic model. Description of haptic metaphors for camera effects.

Different types of movements were designed to explore the potential of haptic feedback for camera effects. The haptic effects have been designed with a home-made editor allowing us to determine the position G'_A of each actuator in time (the *H-Studio*, see Chapter 5). The metaphors were rendered as linear movements for the Arcing, Tilting and Vertigo while more dynamic patterns were used for the other sequences. Moreover with the individual motions of each actuator, we created more complex sensations than the *Cinematic* model.

Figure 6.3 shows the difference between the two models for the Zoom-in sequence. Table 6.2 describes these haptic effects dedicated to the *HapSeat* and what the user is supposed to feel. The implementation of each metaphor is provided in appendix B.

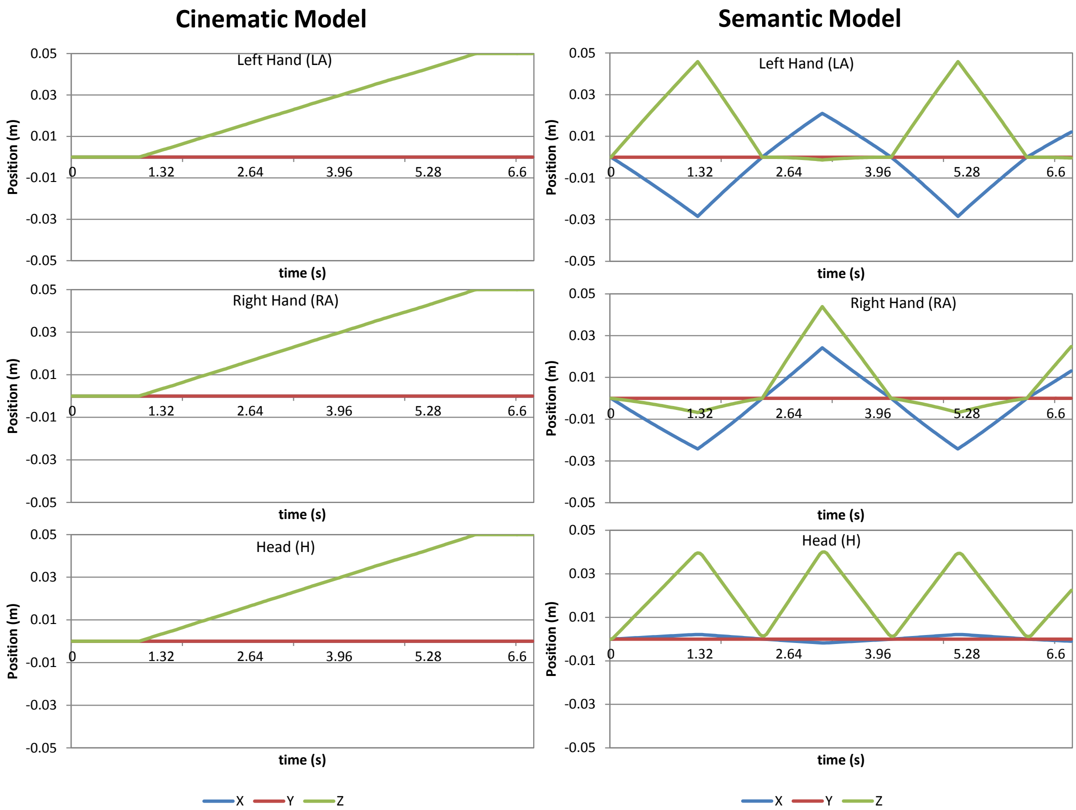


Figure 6.3: Output of the models for the Zoom-in sequence (position of each actuator). With the *Cinematic* model (left), the user feels a global forward movement. With the *Semantic* model (right), a walk forward is simulated. Movements of the left and right actuators are desynchronized.

6.2.5 Haptic rendering

The models provide, for each instant t of the simulation, the target position G'_A (namely G'_H, G'_{LA} and G'_{RA}) for each actuator A (namely H, LA and RA).

Most force-feedback devices (such as the Novint Falcons) are impedance haptic devices, and the position of the actuator is thus not directly controllable. Indeed this kind of device is designed to sense the current position of the actuator and to provide a force feedback to the user. A spring-damper model is thus used to control these devices in pseudo-position. The force $\vec{F_A}$ applied to an actuator A is computed by:

$$\vec{F_A} = k(\vec{G'_A} - \vec{P_A}) - d\vec{V_A} \tag{6.7}$$

where $\vec{G'_A}$ is the targeted position, $\vec{P_A}$ the current position of the actuator, $\vec{V_A}$ its velocity, k the spring constant and d the damping constant.

A haptic-audiovisual player has been developed to play back both video sequences synchronized with haptic feedback. The haptic loop runs at 1KHz and the value of the force $\vec{F_A}$ is updated at each instant t.

6.3 User study

A user study was conducted to evaluate the influence of our haptic effects on the quality of experience (QoE [Jai04, Kil08]), i.e. the subjective user's experience with haptic-audiovisual content. Our hypothesis is that a movie enhanced with our haptic effects provides a better user experience than with a regular movie.

Thirty-eight participants took part in this experiment, aged from 14 to 53 ($\bar{x}$=36.39 σ_x=10.47). Nine were female, 3 left-handed and 9 already used a Novint Falcon. None of them was an expert user of force-feedback devices or motion platforms.

6.3.1 Experimental plan

To evaluate the impact of our models on the QoE, we used four types of haptic feedback.

1. **Cinematic Feedback**: haptic feedback computed using the *Cinematic* model
2. **Semantic Feedback**: haptic feedback computed using the *Semantic* model.
3. **No Haptic Feedback**: only the video was displayed, the actuators remained in the center of their workspace.
4. **Random Feedback**: haptic feedback computed from a low-pass filtered white noise (cutoff frequency $F_c = 0.5Hz$).

The *No Haptic Feedback* corresponds to a regular movie viewing session and serves as a control condition to show how the others feedback modify the QoE. The *Random Feedback*, not synchronized with the video, is used to evaluate the influence of a synchronous feedback on the QoE.

To compare the models we selected a pairwise comparison method: for each video sequence, every feedback was compared against all the others. This led to 6 couples of haptic feedback per sequence (except for the Vertigo where the *Cinematic feedback* is equal to the *No Haptic Feedback*. There were 3 couples in this case). For our 7 sequences, we obtained a total of $6 \times 6 + 3 = 39$ couples (conditions). In order to avoid effect order, the inverse of each couple had also to be tested. Therefore each participant tried 78 conditions.

6.3.2 Procedure

The duration of the study was about 30 minutes, the participant was comfortably installed on the *HapSeat* (see Figure 6.4). The experiment included a training phase in which the participant experienced the seven videos associated to one of the four haptic feedback (randomly chosen). Then the 78 conditions were presented in a random order. Participants were allowed to take a break at any time. For a condition, the participant experienced one video plus an associated haptic effect, then the same video plus a different haptic effect. The requested task was to select the favorite sequence by pressing a button. The next condition was then automatically started. Finally a post-test questionnaire was submitted to collect more information about the user's experience.

Figure 6.4: Experimental Setup, front view (left) and back view (right). The participant experiences haptic effects while watching a video.

The video sequences were made short, seven seconds, to prevent the experiment from being too long and too tiring for the participants. A pilot study was conducted to make sure that the duration of each video sequence was enough to complete the task.

6.3.3 Results

A point was given to a model each time it was chosen by a participant (scores were normalized from 0 to 1 the maximum score). The scores are displayed in Figures 6.5 and 6.6. Scores are denoted by S_X^Y with X for the model and Y for the sequence. The normality of the distributions cannot be assumed according to the Shapiro-Wilk test. Hence non-parametric tests were used to analyze these results: Friedman Anova and Wilcoxon test with Holm-Bonferroni correction (see tables 6.3 and 6.4).

The main result is that the haptic feedback computed from the *Cinematic* model improves the QoE (Friedman Anova: $p < 0.05$). The score for this model is significantly higher than the score for the *None* condition ($S_C^{All} = 0.78 > S_N^{All} = 0.5$, Wilcoxon: $p < 0.05$). The score for the *Random* condition is significantly lower than the others (Wilcoxon: $p < 0.05$) which would mean that a haptic feedback not consistent with the video sequence decreases the QoE. Interestingly the haptic feedback provided by the *Semantic* model is not significantly different from the *None* condition ($S_S^{All} = 0.51 \approx S_N^{All} = 0.5$, Wilcoxon: $p > 0.05$). But this *a priori* equality requires a deeper analysis.

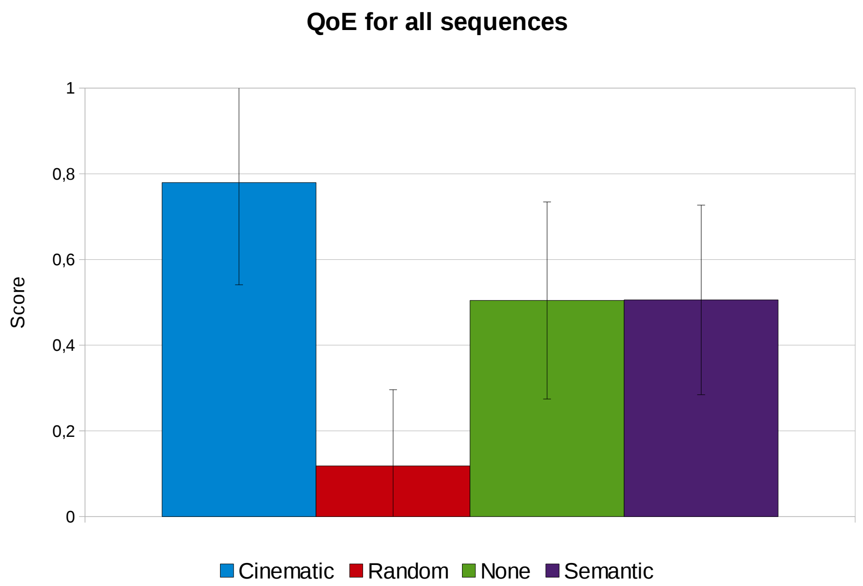


Figure 6.5: Average results for all sequences. The *Cinematic* model improves the quality of experience compared to the *None* condition.

Sequence / **Model**	Arcing	Crane Shot	Dutch Angle	Tilting	Traveling	Vertigo	Zoom	All	
Cinematic	0.82 0.17	0.93 0.11	0.88 0.14	0.75 0.20	0.89 0.12	n/a n/a	0.84 0.20	0.78 0.24	$\bar{x}$ σ_x
Semantic	0.65 0.18	0.34 0.13	0.36 0.15	0.73 0.23	0.43 0.19	0.58 0.11	0.46 0.21	0.51 0.22	$\bar{x}$ σ_x
Random	0.14 0.23	0.13 0.21	0.15 0.20	0.12 0.15	0.13 0.16	0.07 0.14	0.09 0.14	0.12 0.18	$\bar{x}$ σ_x
None	0.40 0.19	0.61 0.21	0.61 0.24	0.40 0.20	0.56 0.20	0.36 0.13	0.61 0.25	0.50 0.23	$\bar{x}$ σ_x
F. Anova	73.9407 3 $6.11e^{-16}$ ***	83.6925 3 $2.2e^{-16}$ ***	78.1257 3 $2.2e^{-16}$ ***	76.0279 3 $2.2e^{-16}$ ***	84.0592 3 $2.2e^{-16}$ ***	57.3099 2 $3.59e^{-13}$ ***	74.2951 3 $5.13e^{-16}$ ***	446.7869 3 $2.2e^{-16}$ ***	χ^2 df p sig.

Table 6.3: Means ($\bar{x}$) and Standard deviations (σ_x) of the score in percent, for each model with respects to each sequence. A Friedman Anova ($\chi^2, df, p.value$) has been performed on each sequence.

The scores for each model and for each sequence are depicted in Figure 6.6. The tendency observed previously is still valid: the score for *Cinematic* model is higher than *None* which is higher than *Random*. Except for the Vertigo sequence where the *Cinematic* model is not applicable in the sense that it provides the same feedback as the *None* condition. Scores for the *Semantic* and *None* conditions are different though. Haptic feedback from the *Semantic* model provides a higher QoE for the Vertigo, Arcing and Tilting sequences (Wilcoxon: $p < 0.05$). For the Tilting sequence, it is not significantly different from the *Cinematic* condition ($S_S^{Ti} = 0.73 \approx S_C^{Ti} = 0.75$, Wilcoxon: $p > 0.05$). Otherwise the score is lower than the *None* conditions for the other sequences (Wilcoxon: $p < 0.05$).

Arcing	Cinematic	None	Random
None	$1.4e^{-10}$	-	-
Random	$3.6e^{-11}$	$1.1e^{-05}$	-
Semantic	0.00011	$1.1e^{-06}$	$1.5e^{-09}$
Dutch A.	Cinematic	None	Random
None	$2.2e^{-06}$	-	-
Random	$6.7e^{-12}$	$6.3e^{-08}$	-
Semantic	$2.4e^{-12}$	$1.8e^{-05}$	$1.8e^{-05}$
Traveling	Cinematic	None	Random
None	$7.6e^{-10}$	-	-
Random	$5.1e^{-13}$	$4.6e^{-09}$	-
Semantic	$2.7e^{-11}$	0.0012	$1.9e^{-08}$
Zoom	Cinematic	None	Random
None	$9.2e^{-06}$	-	-
Random	$1.3e^{-13}$	$9.4e^{-11}$	-
Semantic	$1.0e^{-09}$	0.0044	$2.8e^{-10}$

Crane S.	Cinematic	None	Random
None	$1.4e^{-09}$	-	-
Random	$2.2e^{-12}$	$6.5e^{-08}$	-
Semantic	$1.2e^{-13}$	$3.5e^{-07}$	$2.4e^{-07}$
Tilting	Cinematic	None	Random
None	$1.0e^{-08}$	-	-
Random	$1.3e^{-11}$	$1.0e^{-06}$	-
Semantic	0.83	$1.0e^{-06}$	$1.6e^{-11}$
Vertigo	None	Random	
Random	$2.4e^{-09}$	-	
Semantic	$1.2e^{-09}$	$8.3e^{-13}$	
All	Cinematic	None	Random
None	$< 2e^{-16}$	-	-
Random	$< 2e^{-16}$	$< 2e^{-16}$	-
Semantic	$< 2e^{-16}$	0.61	$< 2e^{-16}$

Table 6.4: Pairwise comparison of each model for each sequence using Wilcoxon test with Holm-Bonferroni correction.

6.4 Discussion

Our results suggest that haptic feedback related to camera effects improves the quality of video viewing experience. Besides, the haptic feedback has to be well-designed otherwise the QoE is decreased such as with the *Random* feedback. Haptic effects directly related to the camera movements (i.e. computed from *Cinematic* model) seem relevant for all sequences while a metaphoric approach manually created with strong hypothesis (i.e. *Semantic* model) is successful for particular cases.

In this study the *Semantic* model was preferred to the *None* condition for three sequences out of seven. The metaphors for these sequences (Arcing, Tilting and Vertigo) were rendered as linear movements while the others were non linear. As the movements of the camera were also linear, we think that the dynamic between the visual stimulus and the haptic feedback is important for users. A huge difference would lead to a feeling of desynchronization. This point may be confirmed by the results of our previous studies (see Chapters 3 and 5): the *Random* feedback was preferred to the *None* feedback with

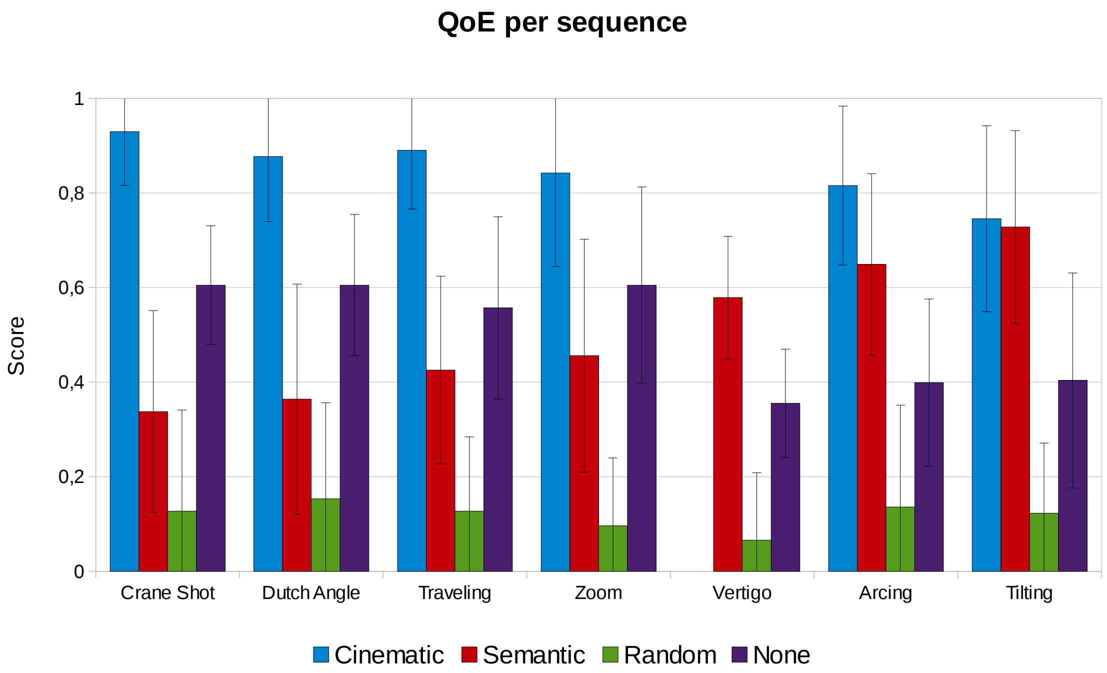


Figure 6.6: Detailed results for all sequences. Score for *Semantic* model is higher than the score for *None* for Vertigo, Arcing and Tilting sequences. Score for the *Cinematic* model is always the highest.

first-person point-of-view video sequences of dynamic events (horse ride, bike ride, car drive). In this case, this feedback was not perceived as totally incoherent.

We have also observed that the direction of the movement of the actuators seems to less impact the QoE. For the Tilting sequence the output of the *Cinematic* model is a backward rotation while the output of the *Semantic* model is a downward movement of all actuators. Directions are different but both were equally appreciated.

Interestingly the metaphors are recognized by several participants. They reported in the post-test questionnaire something similar to a “foot walk” or a “crab walk” for the Zoom-In and Traveling sequences. Some of them even recognized the “flying away” metaphor for the Crane Shot sequence. This would mean that the semantics associated to these effects is understood. However they reported that these haptic effects are not easy to interpret because of the lack of context. According to them, this would work for first-person point-of-view videos or video games where the audience can assume being the main character. Moreover cinematographic effects like the Dutch Angle are designed to be uncomfortable for the user, so the related haptic metaphors are not inclined to be chosen over a *None* feedback.

From these observations we would say that (1) the visual feedback determines the context (dominance of visual over haptic modality). Then (2) the haptic feedback may be perceived as coherent if its dynamic is similar to the visual motion, but (3) it seems unnecessary to follow the same direction. Hence haptic effects should start and stop with the visual stimulus (synchronization) and respect its dynamic (coherence). Deeper investigations are required to determine precise thresholds of the haptic perception in multimedia context, but these results represent a first step in the provision of guidelines for haptic designers.

6.5 Chapter conclusion

In this chapter we exploited the potential of haptics which we considered as a new medium, equivalent to image and sound. We introduced the notion of *Haptic Cinematography* and we proposed a taxonomy of haptic effects for audiovisual content. More precisely a new kind of haptic effects based on cinematographic camera motions was detailed. These cinematographic techniques are extensively used by movie makers to create emotion or ambiance. We believe that haptic feedback can underline these techniques and enhance the video viewing experience.

We proposed two models to render such haptic effects: the *Cinematic* model where parameters of the camera are directly used to make users feel the movement of the camera, and the *Semantic* model based on metaphors reproducing the meaning usually conveyed by the motion of the camera. These two models were implemented on the *HapSeat*.

A user study was conducted to evaluate the relevance of this approach. Results showed that the haptic feedback computed with our models improves the quality of experience while a random haptic feedback decreases it. More precisely the *Cinematic* model is well adapted to all sequences while the *Semantic* model seems effective for specific conditions. In addition, effects should be designed according to the dynamic of the camera movement but the direction of the haptic and camera motions may be different.

Chapter 7

Conclusion

In this Ph.D. manuscript we studied the potential of haptic feedback for enhancing the audiovisual experience. The main goal was to improve the video viewing experience by the stimulation of the haptic modality. We followed two research axes corresponding to two fundamental challenges in the recent field of haptic-audiovisuals (HAV). The first axis (Part I) focused on the **rendering of haptic effects in video viewing settings**. The two objectives of this axis were to propose a new haptic device dedicated to video viewing scenarios, and to adapt haptic rendering algorithms to the haptic-audiovisual experience. The second axis (Part II) focused on the **production of haptic-audiovisual content**. Two objectives were also defined: developing new tools and techniques to enable the creation of haptic effects, and exploring combinations of haptic feedback and audiovisual content in order to propose new haptic effects.

We first **studied and presented the state-of-the-art** in the field of HAV (Chapter 2). The three main challenges of HAV, namely production, distribution and rendering of haptic effects were covered. Existing works related to each challenge were detailed. Besides, techniques and metrics to evaluate the haptic-audiovisual experience were presented. From this review we identified few devices providing a wide range of sensations and suitable for video viewing settings. Research opportunities on the design of haptic effects, as well as on the development of new authoring tools, also appeared.

In the first part of this manuscript we have studied the rendering of haptic effects in video viewing context. We have designed a new device suitable for consumer settings and have developed a new haptic rendering algorithm for haptic-audiovisual content.

To provide haptic feedback in video viewing scenarios, we proposed the *HapSeat*, **a new device to render 6DoF sensation of motion** thanks to three local force-feedback devices (Chapter 3). These actuators, embedded in an armchair structure, apply forces on the user's head and hands mimicking mobile headrest and armrests. We designed two control models to explore different ways to generate sensations of motion with this setup. The Physical model provides the local forces supposed to be felt during a movement, and the Geometrical model reproduces the position and attitude characterizing a movement. A user study showed that the *HapSeat* and both control models succeed in increasing the user's experience in passive navigation scenarios as well as providing a realistic sensation of motion.

Then we focused on the **haptic rendering for haptic-audiovisual content** (Chapter 4). Haptic effects may be designed independently from a specific haptic device, and can happen in a noncontinuous way. To handle the rendering of such effects, we introduce the use of a new **washout filter** for force-feedback devices. We relied on a user's body model to compute kinesthetic perception thresholds. This allows to enhance the haptic rendering and to adapt the haptic feedback to the workspace of the device. A user study was conducted to identify the key parameters in the design of a washout filter. Three profiles were designed, and it appeared that the washout filter should be adjusted depending on the user's preference regarding the synchronization of the effects with the video or their amplitude. The results were generalized by an experiment on an actual short film enhanced with haptic effects.

In the second part of the manuscript we have proposed new tools and techniques to create haptic-audiovisual content. We proposed a new authoring tool offering novel creative perspectives to content creators and we introduced the *Haptic Cinematography* which consider haptics as a medium equivalent to image and sound.

We first introduced the ***H-Studio*, a novel authoring tool which enables the creation of haptic-audiovisual content** (Chapter 5). The tool allows the design of motion effects and their synchronization with a video. Three editing methods were proposed. Two methods take advantage of a force-feedback device to manually edit motion effects: either by setting a direction and an orientation at specific instant of the video or by directly drawing a trajectory. The third method consists in capturing a video and the motion effects. For this purpose we developed a new input device combining a video camera and an inertial measurement unit. Finally, this authoring tool can render motion effects on a force-feedback device, enabling the preview of the effects. A user study showed that the captured motion effects are perceived as realistic and enhance the quality of the audiovisual experience.

Then we explored the **potential of haptic feedback for audiovisuals through the *Haptic Cinematography*** (Chapter 6). We first proposed a taxonomy of haptic effects, and we focused on the coupling of haptic feedback with cinematographic camera motions. We introduced two models to generate such effects. The Cinematic models make the user follow the movements of the camera, and the Semantic model provides haptic metaphors for the camera effects. Results from the user study showed that the direct mapping of the movement of the camera on a haptic device improves the user's experience. Haptic metaphors are also successfully conveyed but need to respect the dynamic of the visual scene to be perceived as coherent.

Future Work

The work presented in this manuscript leaves some questions unanswered, which could be addressed in short-term future work. We present future research possibilities according to our four objectives presented in the introduction of this manuscript.

New haptic device dedicated to video viewing settings

- **Prototype.** The prototype of the *HapSeat* stimulates the user's head and hands. It would be interesting to add more points of stimulation in order to increase the user's immersion. The stimulation of the feet or the legs could significantly improve the setup. In the current configuration the user's feet touch the ground which may be contradictory with a motion effect. Also, a lack of feedback could be felt between the force-feedback devices, reducing the sensation of a global effect of motion. Small actuators could fill the gap between these devices, making the haptic feedback more united. The addition of vibrating motors, in the back for instance, would also be an interesting enhancement of the setup. The *HapSeat* could be combined with the seat designed by Israr et al. for instance [IP11].

- **Control Models.** The models provide the same haptic feedback for the head and the hands (in terms of amplitude). Results from the user studies pointed out that the rendering applied to the head has to be managed differently than for the hands. The movement of the actuator does not need to be large to be perceived. Moreover vibrations applied to the head decrease the comfort of the setup. Such effects are really immersive but have to be limited to the hands. The Physical model could also be improved by relying on a human body model instead of a rigid body model. The forces computed would thus be more realistic.

- **Evaluation.** Further evaluations could be conducted to finely characterize the simulation of motion with the *HapSeat*. Simulation providing 6DoF motion effects should be used to explore the full potential of the setup. Besides a comparison to a classical motion simulator could be useful. Even if the *HapSeat* is not designed to provide a strong sensation of motion, it would be interesting to identify to what extent it could replace a motion platform.

New haptic rendering algorithm for haptic-audiovisual scenarios

- **Optimization of the washout filter.** Three profiles have been defined to tune the washout filter. One focuses on the synchronization between the effects and the video while the others try to preserve the amplitude of the effects. The washout filter may be improved in order to limit the trade-off between synchronization and amplitude. The global scaling performed by the algorithm could be replaced by a more local and dynamic scaling. Besides user studies are required to evaluate how users perceive the difference between amplitudes of movements. It may not be necessary to keep the exact amplitude of the effects.

- **Cognitive washout.** Further studies are required to understand the integration of haptic-audiovisual stimuli by the user. While watching a video the user's attention might be focused on the screen, and therefore the kinesthetic perception may be less sensitive. Hence the washout filter could be performed at a cognitive level rather than a pure haptic level.

New authoring tool for creating of haptic effects

- **Usability studies.** Three methods for designing motion effects were proposed: two methods relying on a force-feedback device to manually edit effects and one method based on the import of motion effects from a capture device. Only this last method was evaluated. A user study of the two others methods would also be necessary to identify the usage for which one would be better than another. Besides user studies should be conducted with VFX artists who may be the future users of such a tool.

- **Automatic Extraction.** The automatic extraction of haptic effects has been quickly addressed with the generation of vibration effects from the audio track of the video. This feature could be adapted to the generation of motion effects from the visual content [Tho06] or metadata (the MPEG-7 format includes information about the camera [CSP02]). Such a feature could help the content creator to quickly prototype a motion effect which could then be adjusted.

New haptic effects for enriching the haptic-audiovisual experience

- **Exploring the taxonomy of haptic effects.** The coupling of haptic feedback and cinematographic camera effects was studied in details. But others effects were proposed in the taxonomy (related to the montage, the music, etc.). Deeper investigations could then be conducted to evaluate those effects.

- **Combination of diegetic and non-diegetic effects.** The studies in this manuscript focused on the use of diegetic or non-diegetic effects. A combination of these two types, as it is already done for the sound in movie, could be worthy of study. But further investigations are required to understand how such effects can be combined and what would be the impact on the user's experience.

Long-Term Perspectives

In addition to the short-term future work mentioned above, this Ph.D. thesis also paves the way for new research directions and long-term views. Some of these aspects are described below.

Production of haptic effects

The authoring tool presented in this manuscript focuses on the edition of vibration and motion effects. But the range of haptic sensation is much more wider, and many others effects may be included in such a haptic editor (pressure, temperature, etc.). However the ideal editor cannot be a simple extension of the current approaches where each haptic effect is represented by a track ([WRTH13, Kim13] and the *H-Studio*). The edition of complex haptic sensation would be complex. Future research could focus on the design of rich haptic sensations, potentially located on multiple part of the user's body.

In line with the edition of numerous haptic sensations, new capture devices may be designed to record haptic effects during the shooting of the audiovisual content. For example, data related to the temperature or the wind direction in the scene could be recorded. Actors could also be equipped with pressure sensors. Data would then be used to recreate the ambiance during video viewing and to increase the immersion of the audience. With haptics considered as an actual medium, shooting a movie would mean capturing images, sound and also haptic information.

More generally the edition of haptic effects should be integrated in the process of movie making. From the shooting to the post-production. This could lead to the new professional activity of "haptographers" which may be seen as an equivalent of the existing "stereographers" specialized in 3D for movies. To reach this goal, research on HAV should be conducted in parallel to research in cinematography. Nevertheless, HAV is not limited to the video viewing context. Many other entertainment applications could benefits from the contributions in this field of study. Obviously video games could directly use the results but this may open new perspectives for education, tele-learning, tele-contact, medical simulation, etc. [EOEC11].

Distribution of haptic effects

The issue of distributing haptic effects was not addressed in this manuscript. At the time of starting this Ph.D. thesis, the MPEG group was formalizing the MPEG-V, a standard defining sensorial effects (haptic but also visual and olfactory effects) for audiovisual content. Such a standard is necessary to democratize and distribute videos enhanced with haptic effects. Results from the new field of study of HAV will probably highlight the limits of this young standard and also contribute to its evolution. For example the concept of haptic metaphors presented in Chapter 6 is hardly compatible with this format. Yet it would be interesting to describe such high-level effects.

The main challenge in the formalizing of haptic effects is to describe haptic sensations independently from any device, while providing enough information to enable this sensation to be generated by a mechanical device. Research has to be conducted to map haptic sensations to haptic stimuli. For example Obrist et al. have linked tactile experiences to vibrotactile stiumulii [OSS13]. Such results could be useful to improve the MPEG-V or to design other standards.

Rendering of haptic effects

Research perspectives in the rendering of haptic effects can be seen from hardware and software point-of-views. Today haptic hardware provides only one type of sensation (force, vibration, pressure, etc.), localized on a specific part of the body (usually the hand). Yet a haptic experience is a full-body experience resulting from multiple sensations at a time. The potential of the sense of touch is thus not fully exploited. Research on the haptic perception is required as well as research in mechanical engineering to enable the development of new devices.

On the software side, haptic rendering algorithms should handle the variety of haptic devices. The work presented in this manuscript focused on one type of video viewing context, where the user is comfortably seated (home cinema or movie theater). But movies are now consumed on TV, computers, tablets or even mobile phones. The video is already adapted to the screen resolution to optimize the user experience [CLM08]. In a same way, haptic rendering could adapt the generation of haptic effects according to the devices available. A motion effect will be then rendered differently if the video is watched in a 4D theater equipped with motion platforms, at home on a *HapSeat* or on a mobile phone embedding a vibration motor.

Evaluation of the quality of experience

In this manuscript, the quality of experience was systematically evaluated for every contribution proposed. It appeared however that the QoE with haptic-audiovisuals is difficult to characterize and there is a lack in the literature on this new topic. The QoE is mostly evaluated through questionnaires which are relevant for collecting the subjective user's experience. The capture of physiological data could be an interesting technique to collect a more objective measure.

In a first approach to evaluate the QoE, we identified several components of the QoE (Realism, Sensory, Comfort and Satisfaction). User studies are needed to evaluate these factors. The next step would be to build a model of the haptic-audiovisual experience. Such an approach is proposed by Hamam et al. through a taxonomy of items composing the QoE with haptic applications [HESG08]. Furthers studies are required to validate these factors and to determine how much each of them contribute to the QoE.

Such a model should not be limited to the simple addition of haptics to audiovisuals though. Research is currently conducted on the evaluation of the quality of the video viewing experience augmented with other cues such as 3D [HTBLC11] or sensorial effects [WT10]. Eventually the model of the QoE should include all these effects. But there is still a lot to do in each field of study to understand how each individual cue impacts the QoE. Therefore merging of all the effects in order to design a complete model of the quality of the video viewing experience will probably bring new and interesting challenges.

A lot of work remains to be done in order to make HAV a mature technology. Nevertheless the recent research results and technology developments assess the growing interest in this new field of study. There is no doubt that haptics has the potential to enhance the audiovisual experience and will be used to create more and more immersive applications. We hope that the work presented in this manuscript is a first step along this ambitious path.

Appendix A

Haptic Effects Used in Chapter 4 (Washout Filter)

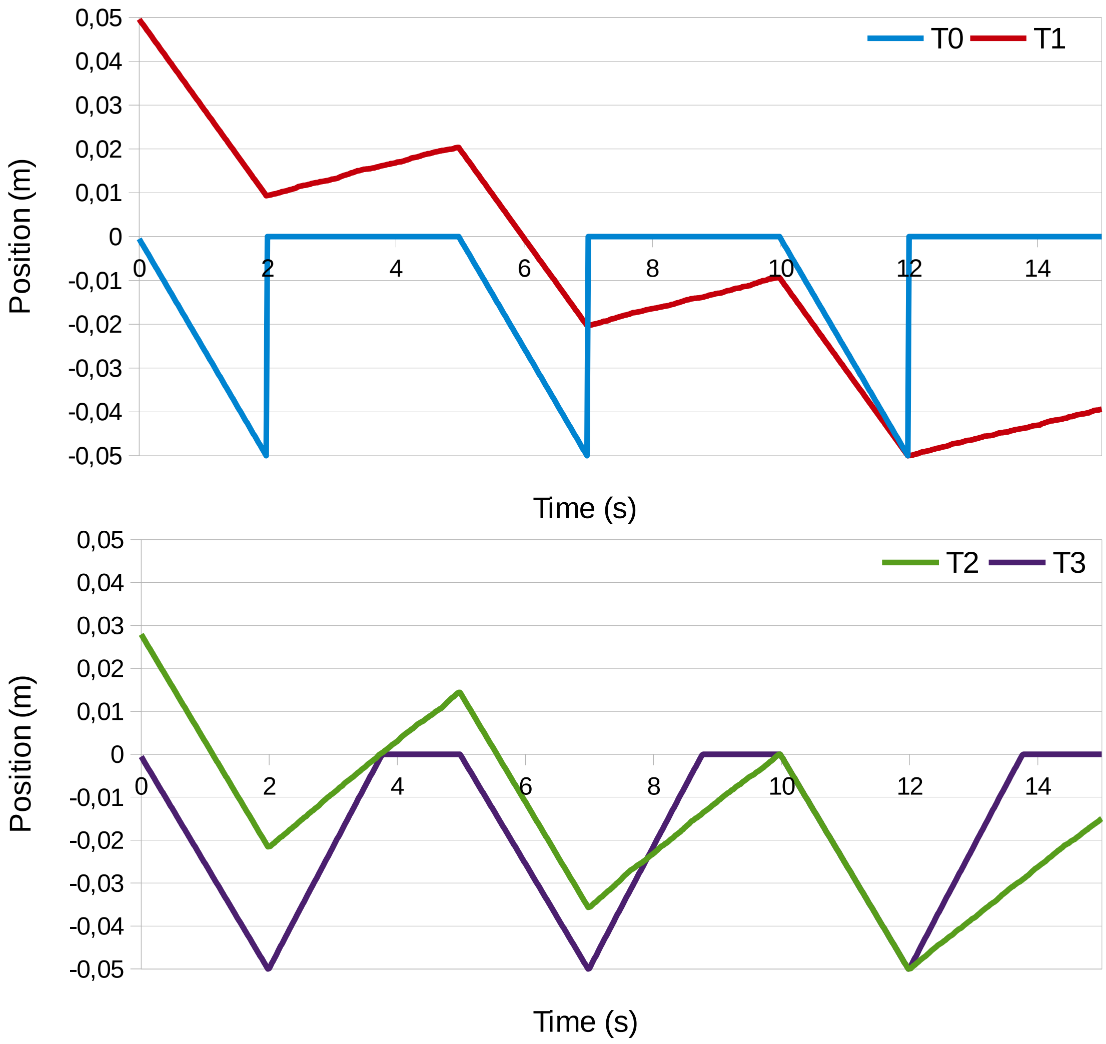


Figure A.1: Haptic effects for the sequence S1. The curves show the positions of an actuator according to the profile selected (T0, T1, T2 or T3). Three effects have been designed here.

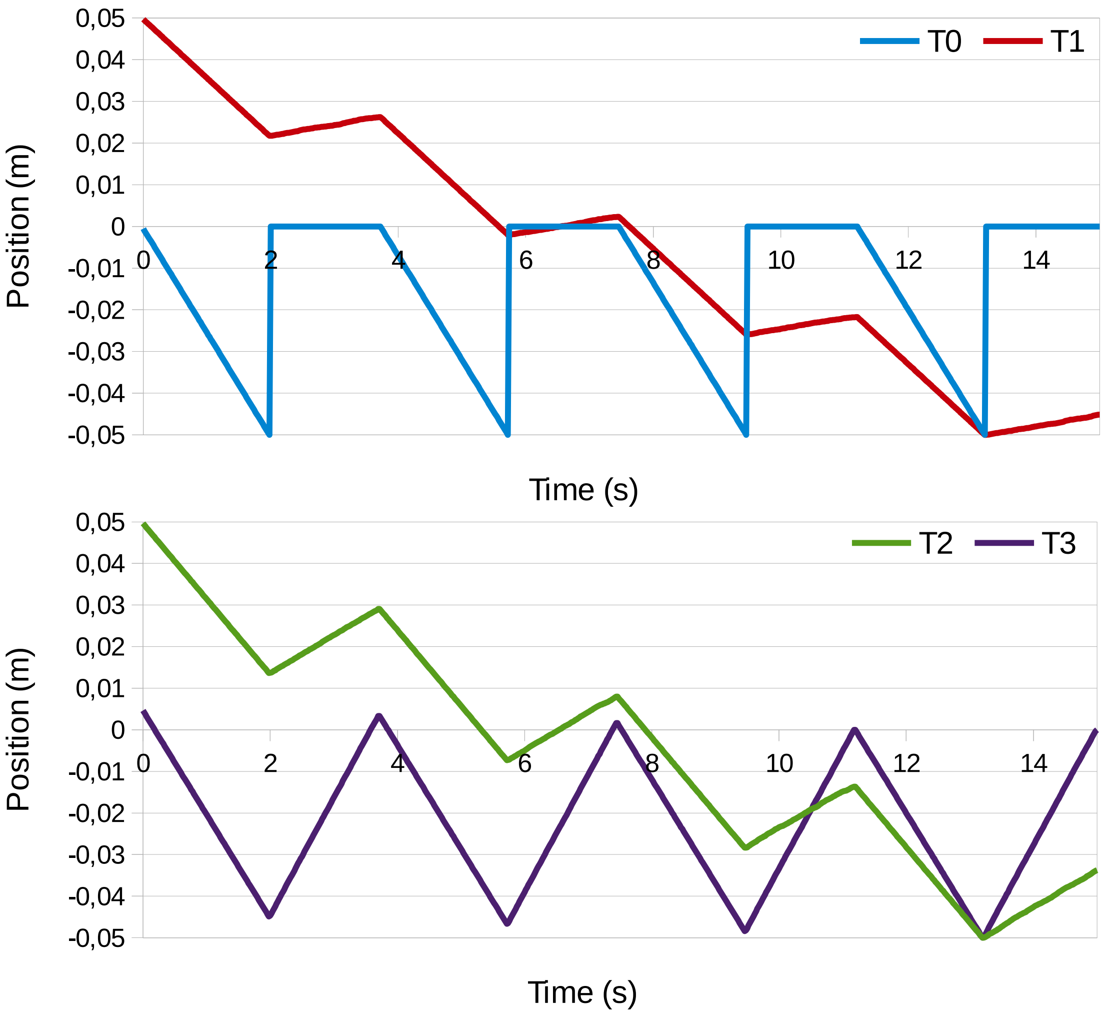


Figure A.2: Haptic effects for the sequence S2. The curves show the positions of an actuator according to the profile selected (T0, T1, T2 or T3). Four effects have been designed.

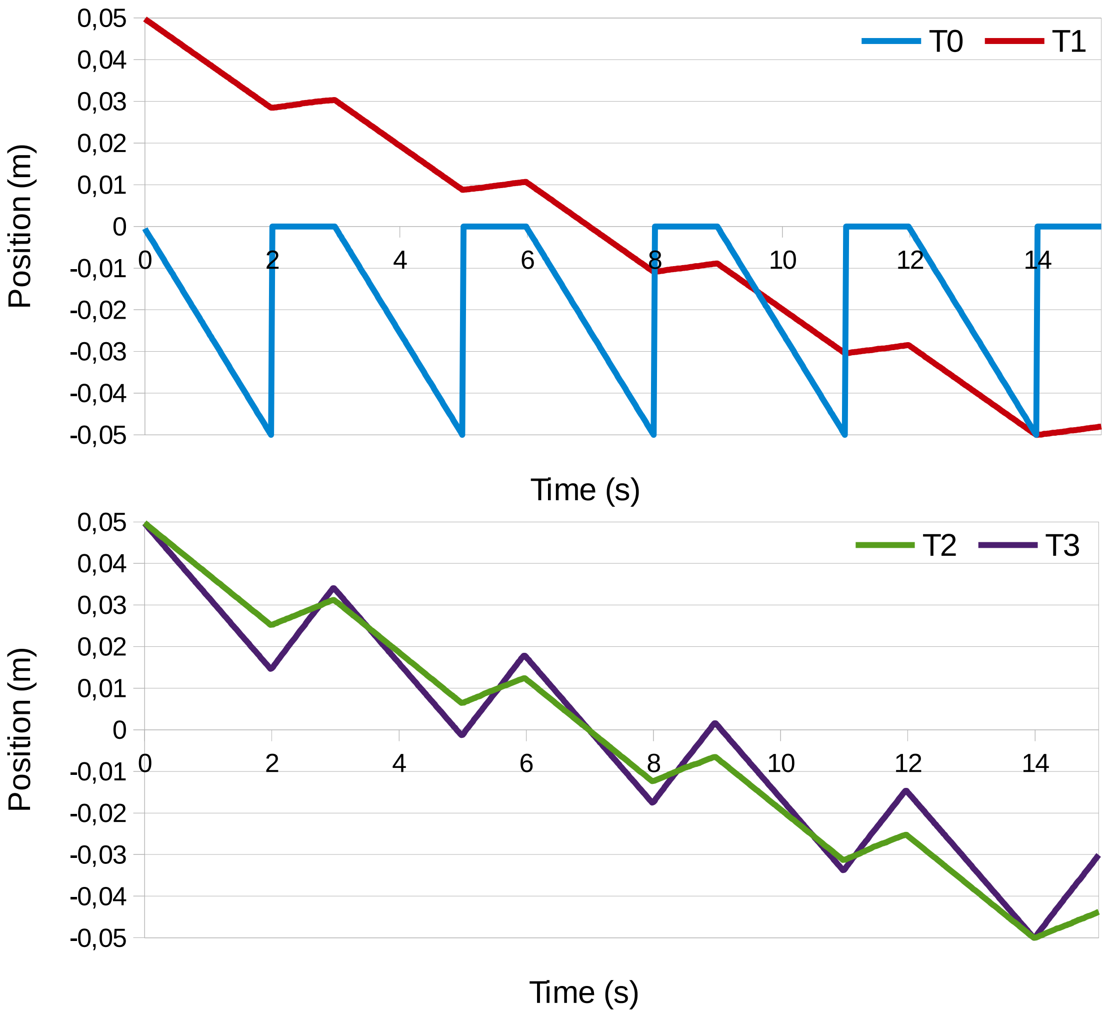


Figure A.3: Haptic effects for the sequence S3. The curves show the positions of an actuator according to the profile selected (T0, T1, T2 or T3). Five effects have been designed.

Appendix B

Output of Cinematic and Semantic models

Output of the two models for each sequence (*Cinematic* on the left, *Semantic* on the right). The position in meters is plotted for each actuator LA, RA and H, and for each axis.

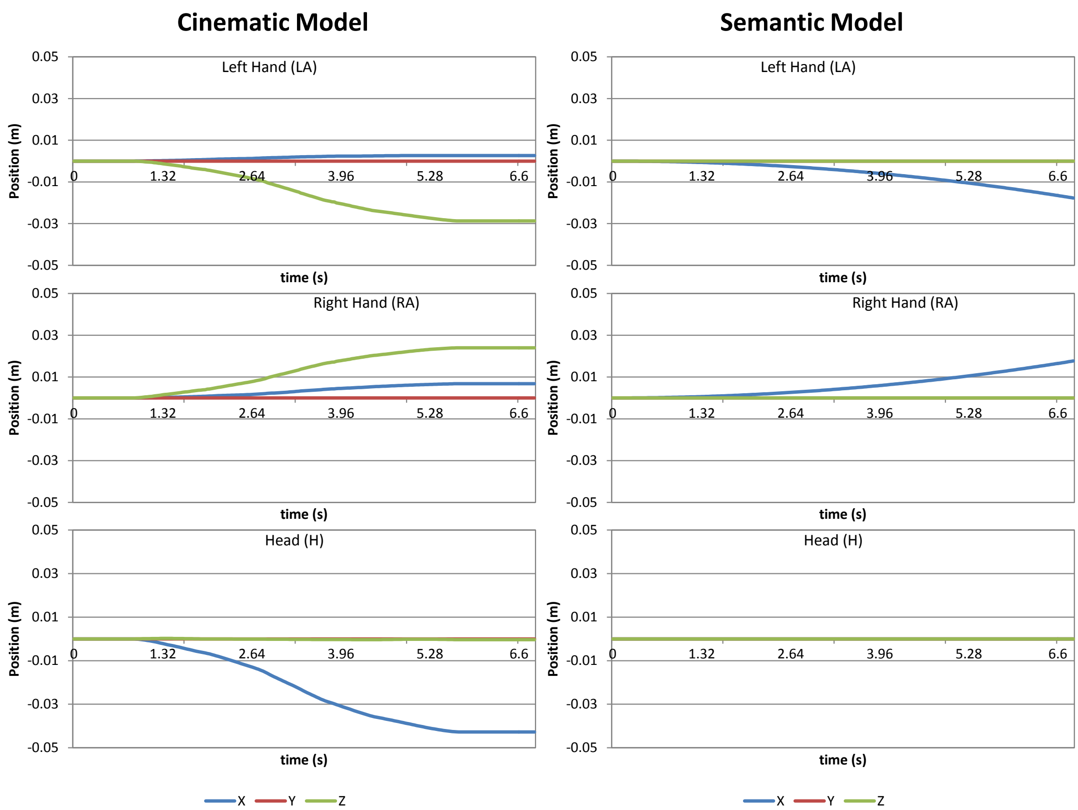


Figure B.1: Arcing

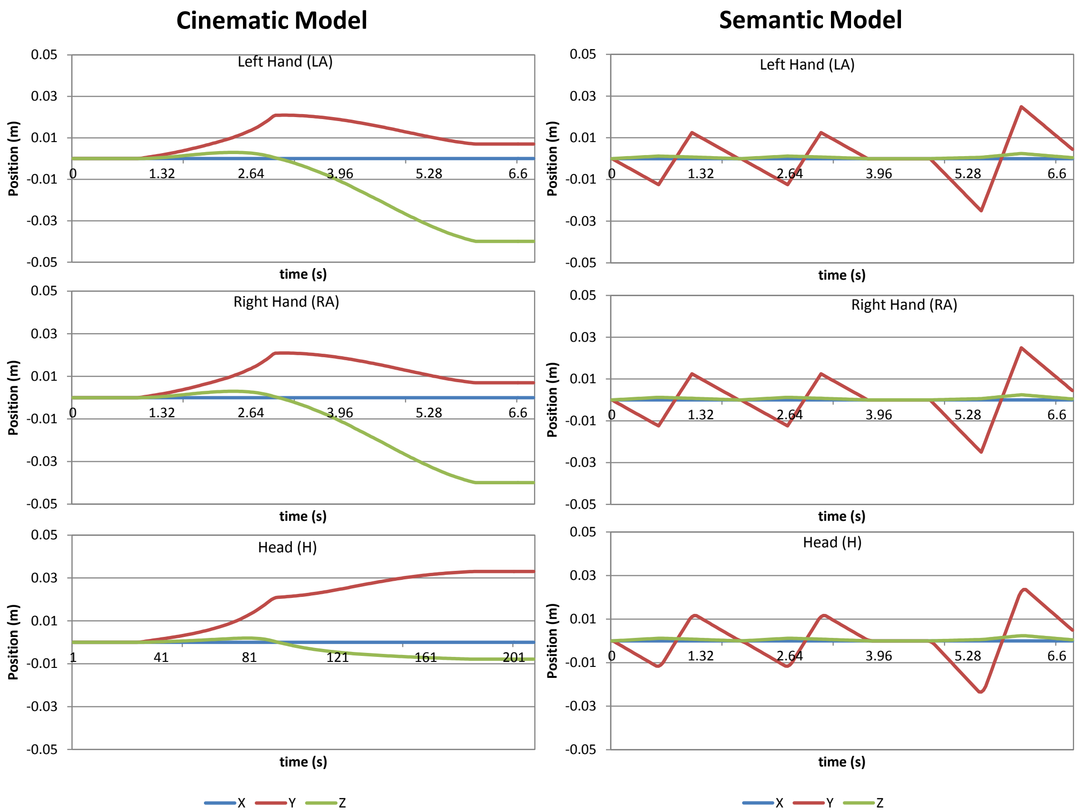


Figure B.2: Crane Shot

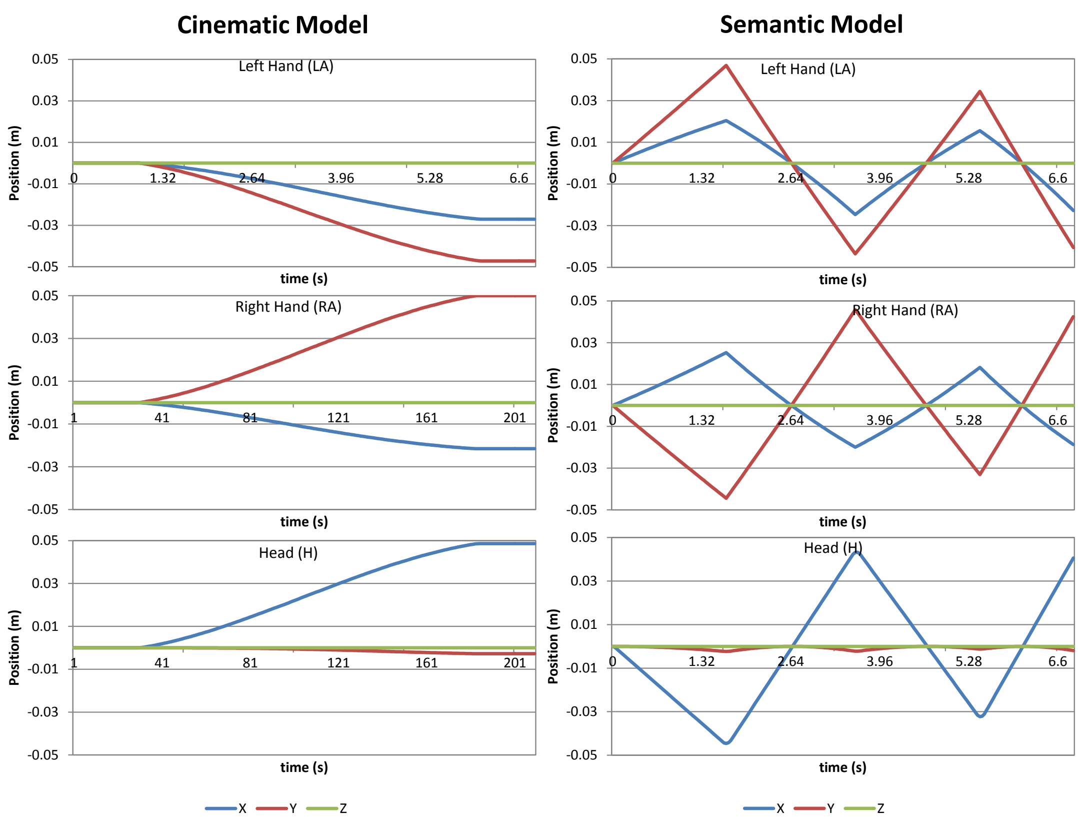


Figure B.3: Dutch Angle

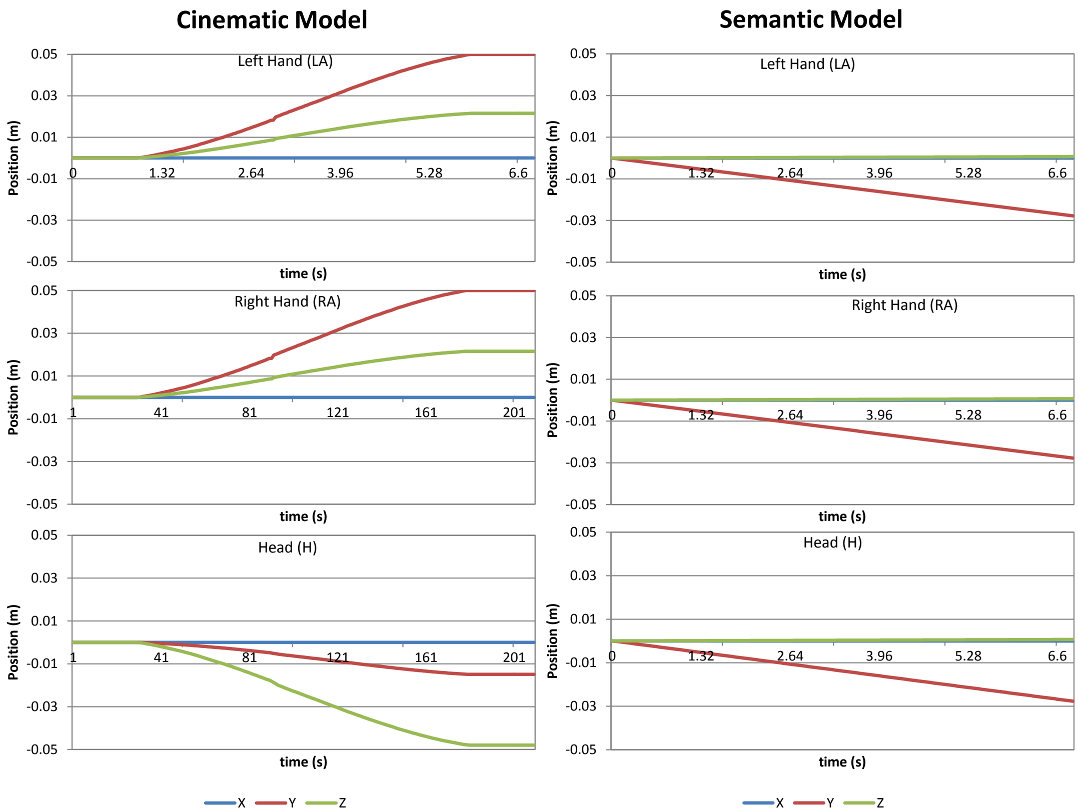


Figure B.4: Tilting

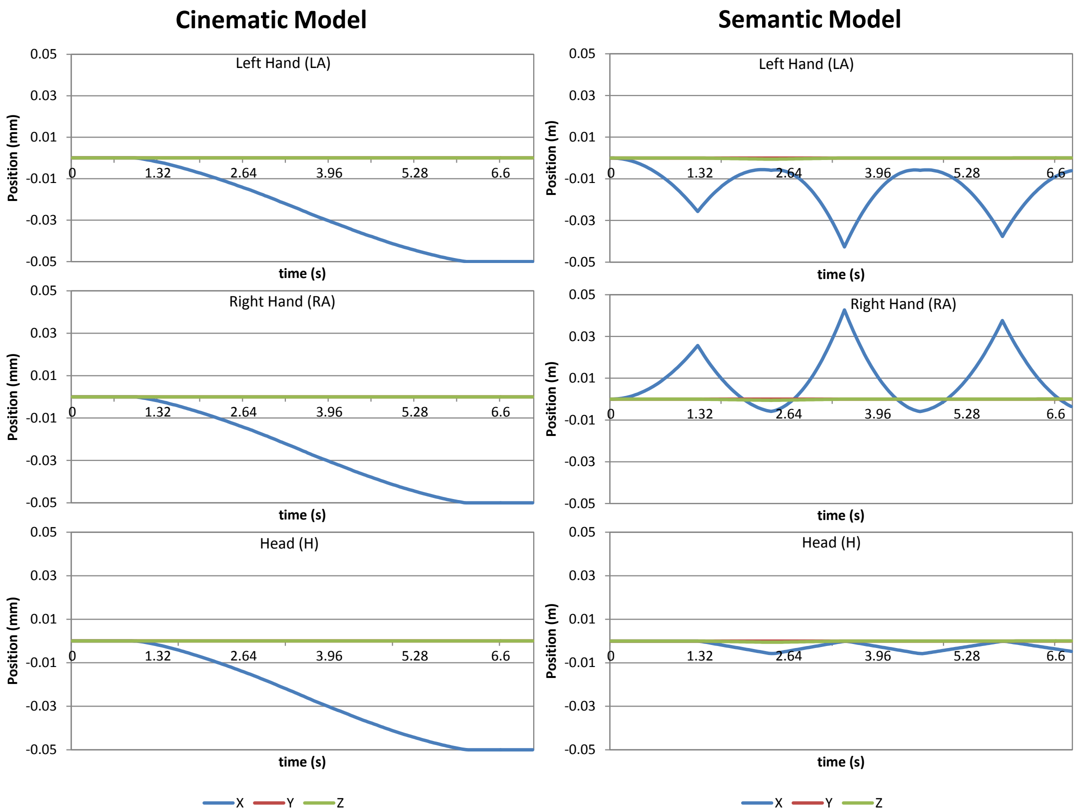


Figure B.5: Traveling

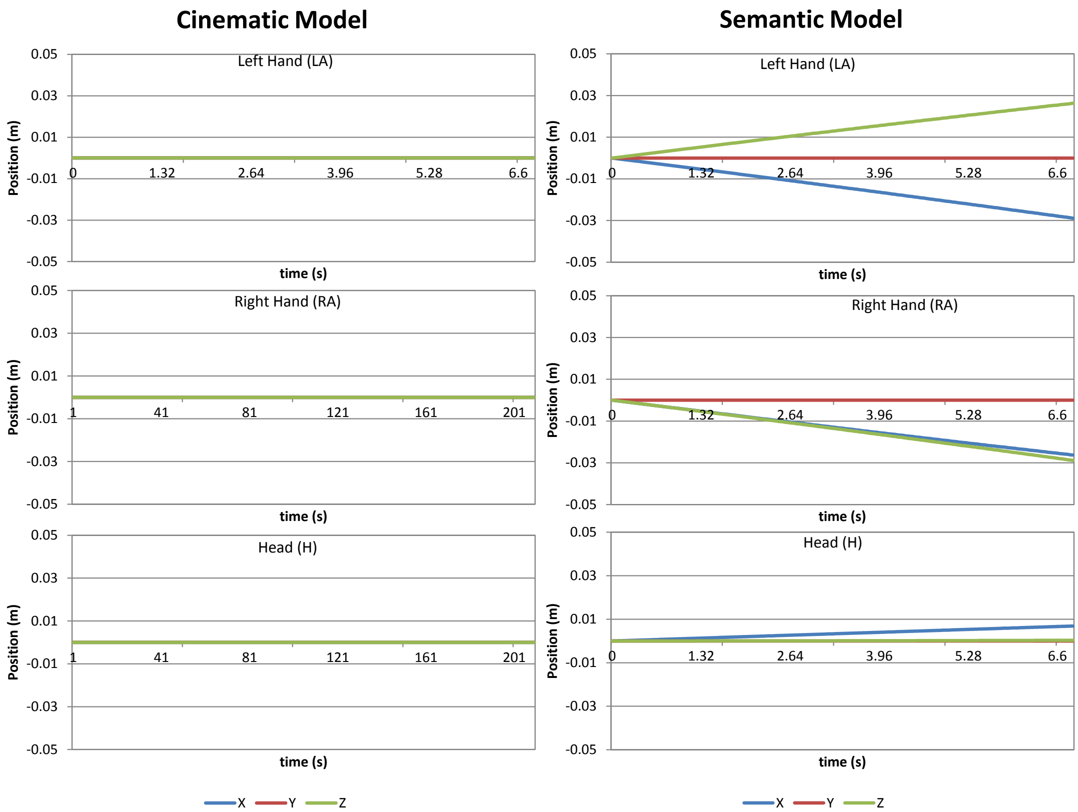


Figure B.6: Vertigo

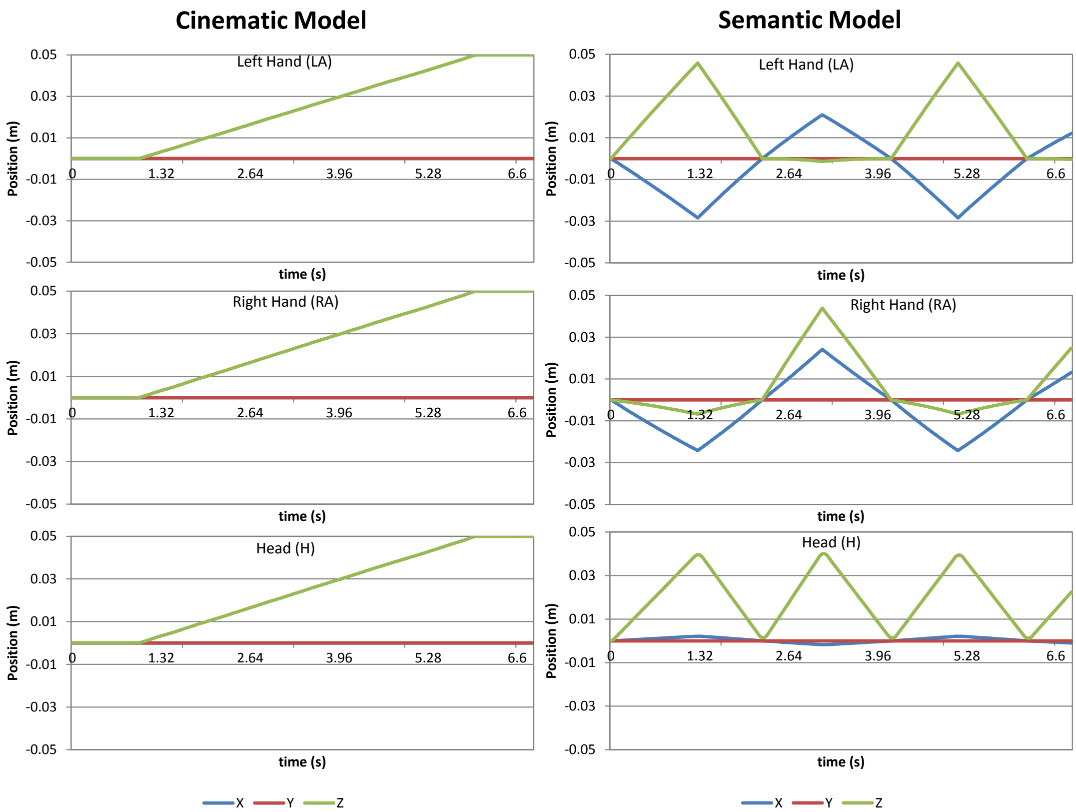


Figure B.7: Zoom

Appendix C

Video Sequences Illustrating the Camera Effects

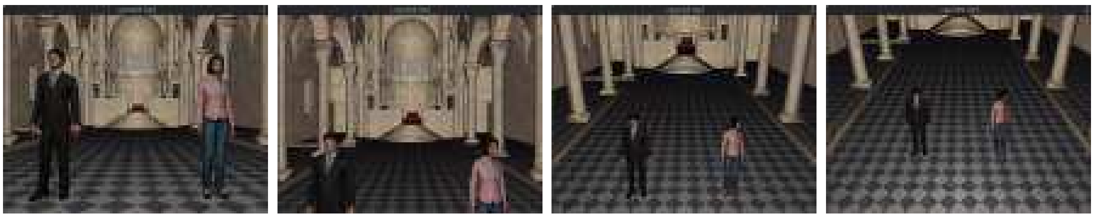


Figure C.1: Screenshots of the Crane Shot sequence.

Figure C.2: Screenshots of the Arcing sequence.

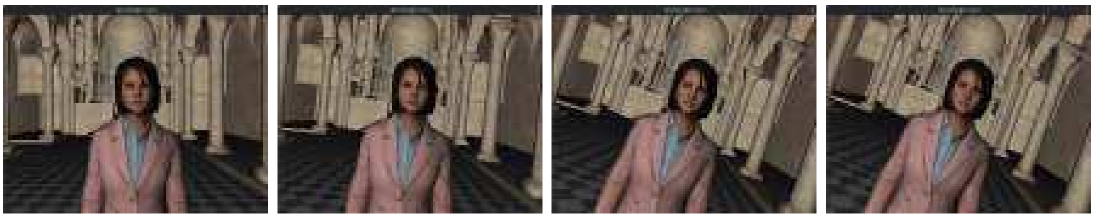


Figure C.3: Screenshots of the Dutch Angle sequence.

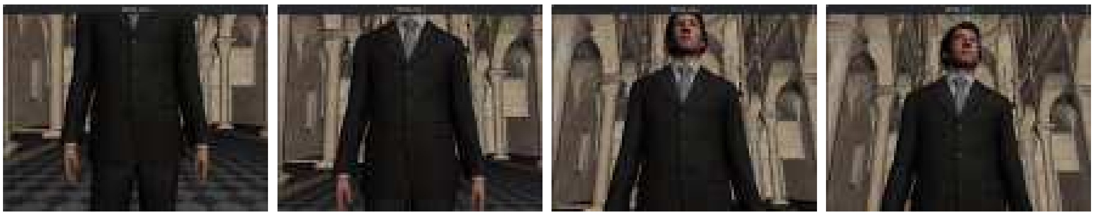


Figure C.4: Screenshots of the Tilting sequence.

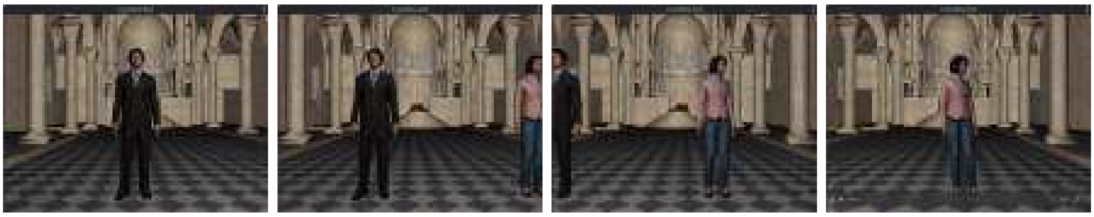


Figure C.5: Screenshots of the Traveling sequence.

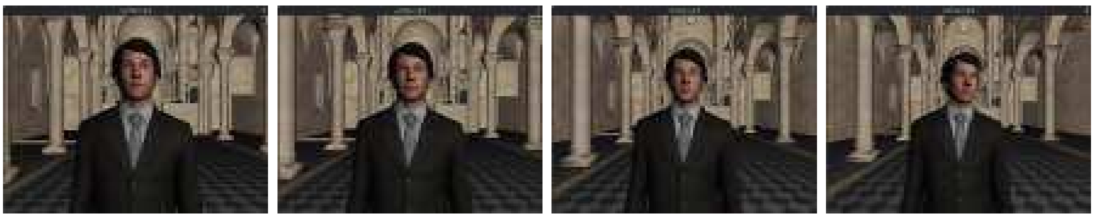


Figure C.6: Screenshots of the Vertigo sequence.

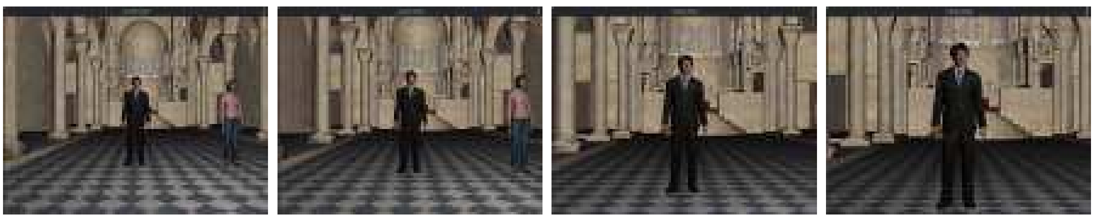


Figure C.7: Screenshots of the Zoom sequence.

Appendix D

Résumé Long en Français

L'importance du sens du toucher (sens haptique) a été particulièrement étudié et apparait être un facteur clé pour l'immersion de l'utilisateur dans les systèmes interactifs. De nombreuses interfaces haptiques permettant l'interaction physique avec des objets distants ou virtuels ont été développées et étudiées. De nos jours les technologies haptiques sont utilisées dans de nombreuses applications médicales, robotiques ou encore artistiques.

Les interfaces haptiques sont au contraire peu employées dans les applications multimédia. Pourtant, en 1962 Heilig a introduit le *Sensorama*, un système où l'on pouvait voir un film en 3D, ressentir des vibrations, du vent et sentir des odeurs [Hei62]. Malgré un fort potentiel pour l'industrie cinématographique, la recherche et les développements technologiques pour l'audiovisuel se sont focalisés sur l'amélioration de l'image et du son. Peu de systèmes, tel les "cinémas 4D", exploitent actuellement les technologies haptiques. Cependant le nombre d'articles mettant en avant le potentiel de ces technologies pour le multimédia est en constante augmentation. O'Modhrain et al. ont démontré que les bénéfices observés des interfaces haptiques dans les systèmes de réalité virtuelle, de téléopération ou dans les jeux vidéo sont transférables aux applications multimédia [OO03]. Les retours haptiques peuvent améliorer les sensations de réalisme, d'immersion, et l'engagement de l'utilisateur dans le contenu [MTB06]. Ils peuvent également représenter plus que des évènements physiques et pourraient transmettre de l'information ou susciter de l'émotion. Ainsi, la combinaison de retours haptiques et de contenus audiovisuels tend vers un nouveau medium, l'haptique-audiovisuel (HAV [EOEC11]), avec ses défis scientifiques qui lui sont propres.

Ce jeune champ d'étude introduit de nouvelles questions. Comment un retour haptique peut-il être combiné efficacement avec des images et du son, et comment ces retours peuvent-ils être conçus? Quel type d'appareil est adapté pour le rendu de retours haptiques dans un contexte cinématographique (cinéma ou domicile de l'utilisateur, potentiellement partagé)? De plus, dans quelle mesure les interfaces haptiques peuvent-elles influencer l'expérience audiovisuelle, et comment la qualité de cette expérience peut-elle être évaluée?

Combiner des retours haptiques et des contenus audiovisuels: défis et contexte

Les grands défis pour combiner des retours haptiques et des contenus audiovisuels peuvent être organisés en un processus de trois étapes: production, distribution et rendu d'effets haptiques (voir Figure D.1). Le terme "effet haptique" est employé pour désigner l'utilisation de retours haptiques avec un contenu audiovisuel [OO03, YAMS06, CES09].

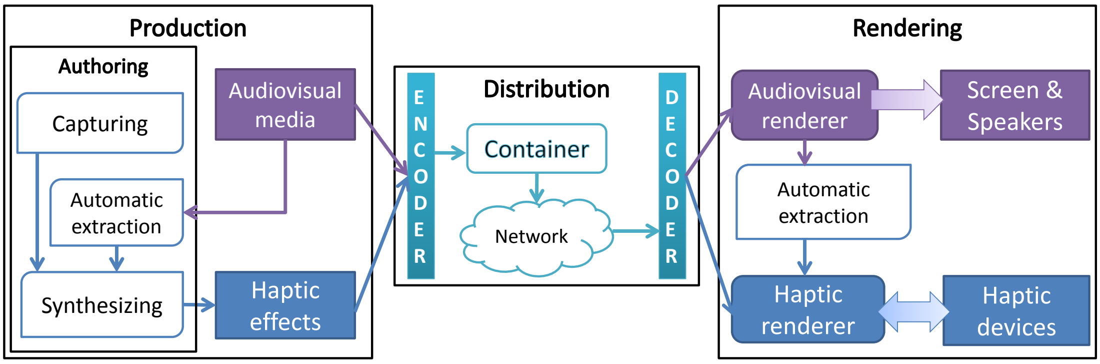


Figure D.1: Processus pour combiner des effets haptiques à des contenus audiovisuels. Les effets sont produits, distribués et rendus en parallèle au contenu audiovisuel.

La première étape consiste à produire le contenu, c'est à dire à créer ou générer des effets haptiques synchronisés avec le contenu audiovisuel. Trois techniques sont décrites dans la littérature: la capture et le traitement de données acquises par des senseurs, l'extraction automatique depuis une composante du contenu audiovisuel (image, son ou annotations), et l'édition manuelle.

La deuxième étape est la distribution des effets haptiques. Les technologies actuelles permettent la distribution de masse par les réseaux. Les retours haptiques se doivent d'être compatibles avec ces technologies, ce qui soulève la question de la formalisation des effets haptiques.

Enfin la troisième étape réside dans le rendu des retours haptiques. Un appareil spécifique doit être utilisé pour que l'utilisateur ressente les effets. Des algorithmes de rendu haptiques sont également employés pour convertir les effets haptiques en commandes pour ces appareils.

Un dernier aspect à considérer, complémentaire au processus et transverse à ses trois étapes, est l'évaluation de la qualité de l'expérience de l'utilisateur (QoE). Cette QoE peut avoir plusieurs définitions [Jai04, Kil08], mais dans notre contexte est décrite comme l'expérience subjective de l'utilisateur avec un contenu audiovisuel.

Objectifs et approche

Ce manuscrit est divisé en deux parties, chacune correspondant à un axe de recherche. La première se focalise sur le rendu des effets haptiques et la seconde sur leur production.

Dans le premier axe nous nous focalisons sur le rendu d'effets haptiques dans un contexte cinématographique. Peu d'appareils existent pour générer des effets haptiques lors du visionnage d'une vidéo, et ils sont soit encombrants et chers (simulateurs de mouvement) ou soit limitants par leur panel d'effets réduit (tablettes ou manettes avec vibreur). Ainsi le premier objectif de cet axe est de proposer un nouvel appareil haptique dédié à un usage cinématographique et générant des effets haptiques immersifs.

L'utilisation d'interfaces haptiques dans un contexte audiovisuel entraine aussi de nouveaux problèmes au niveau des algorithmes de rendu. Les effets haptiques peuvent être créés indépendamment des interfaces et doivent donc être adaptés aux contraintes de l'appareil utilisé par ces algorithmes. Le deuxième objectif de cet axe est alors de proposer un nouvel algorithme de rendu haptique prenant en compte ce problème.

Dans le deuxième axe de recherche nous nous intéressons à la production d'effets haptiques. Jusqu'à lors, ceux-ci sont souvent conçu par les constructeurs d'appareils pour "cinémas 4D". Ils ne sont pas réalisés durant la création des contenus audiovisuels et ne sont donc pas vraiment une composante à part entière de ces média. Cependant peu d'outils permettent la création de tels contenus. Le premier objectif de cet axe est de proposer un nouvel outil de création de contenus haptique-audiovisuels.

Il apparait également que les effets haptiques sont souvent utilisés pour représenter des évènements physiques (explosions, accélérations, etc.). Pourtant les retours haptiques pourraient transmettre de l'information ou susciter des émotions. Le deuxième objectif de cet axe est d'explorer les combinaisons haptiques-audiovisuelles et de proposer de nouveaux effets haptiques.

Nos contributions sont détaillées par la suite, suivant les deux axes de recherche mentionnés précédemment. Nous avons suivi une approche centrée utilisateur tout au long de ces travaux, et nous avons conçu des métriques pour évaluer la qualité de l'expérience. Nos contributions ont été systématiquement évaluées dans cette thèse.

D.1 Partie I - Rendu d'effets haptiques: nouvel appareil et algorithmes pour rendre des effets haptiques dans un contexte de cinéma

Dans cette première partie de la thèse nous nous focalisons sur le rendu d'effets haptiques. Nous proposons premièrement un nouvel appareil, le *HapSeat*, puis deuxièmement un nouvel algorithme de rendu.

D.1.1 HapSeat: simulation de sensations de mouvement avec plusieurs appareils à retour de force intégrés dans un siège

Les simulateurs de mouvement sont traditionnellement basés sur des plateformes de Stewart [Das00]: une plateforme mobile grâce à six cylindres hydrauliques. Typiquement le corps entier de l'utilisateur est mis en mouvement pour simuler des sensations comme des accélérations, des chutes ou des bosses. Ces appareils ne sont pas conçus pour un usage domestique et sont assez chers pour le marché de masse. Ainsi les expériences immersives avec des effets de mouvement sont encore limitées aux parcs d'attractions ou aux "cinémas 4D".

D.1.1.1 HapSeat

Nous proposons un nouvel appareil pour enrichir l'expérience audiovisuelle avec des effets de mouvement à 6 degrés de liberté (DDL). Au lieu de bouger tout le corps de l'utilisateur comme cela est fait traditionnellement avec les simulateurs de mouvements, seulement trois parties du corps de l'utilisateur sont stimulées: les mains et la tête. La perception du mouvement résulte de la stimulation de plusieurs parties du corps (système vestibulaire, organes viscéraux et système kinesthésique [Ber00]). Notre hypothèse est que les stimulations haptiques locales suffisent à générer des sensations de mouvement.

D.1.1.2 Implémentation

Pour illustrer notre concept nous avons développé un prototype utilisant trois bras à retour de force (Novint Falcons [NOV]). Ces appareils sont intégrés dans une structure en aluminium en forme de siège (voir Figure D.2).

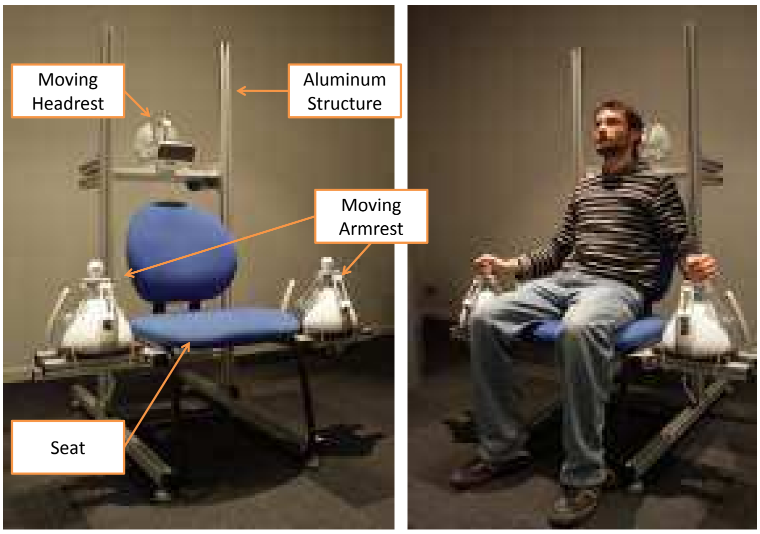


Figure D.2: Prototype du HapSeat. Gauche: structure du siège intégrant 3 bras à retour de force. Droite: système en utilisation.

Nous avons également développé deux modèles de contrôle pour le *HapSeat*. Le premier modèle, dit modèle *Physique*, a pour but de faire ressentir à l'utilisateur les forces subies lors d'un mouvement. Par exemple, si une trajectoire dans un rond point est simulée, l'utilisateur ressentira l'effet de la force centrifuge. Les bras à retour de force iront vers la droite, comme étant repoussés par le rond point. Le deuxième modèle, dit modèle *Géométrique*, reproduit la position qu'aurait l'utilisateur lors d'un mouvement. Pour reprendre l'exemple du rond point, l'utilisateur aura l'impression de tourner vers la gauche. Le bras à retour de force à gauche ira vers l'arrière tandis que celui à droite ira vers l'avant.

D.1.1.3 Evaluation

Afin d'évaluer le mouvement simulé par le *HapSeat* et son influence sur la qualité de l'expérience audiovisuelle, nous avons conduit une étude utilisateur. 17 participants ont pris part à cette étude. Ils devaient regarder deux vidéos montrant un trajet en voiture du point du vue du passager. Pour chacune des vidéos, 4 types de retour haptique étaient proposés: un retour calculé par le modèle *Physique*, un par le modèle *Géométrique*, un retour aléatoire (les bras haptiques bougeaient sans être synchronisés au contenu audiovisuel), et un retour nul (les bras haptiques ne bougeaient pas). La qualité de l'expérience était évaluée par un questionnaire.

L'étude a montré que le retour haptique produit par le *HapSeat* améliore la qualité de l'expérience. L'utilisateur préfère une séquence audiovisuelle avec un retour haptique plutôt qu'une séquence classique. Aussi la synchronisation des retours haptiques avec l'audiovisuel est nécessaire (un retour haptique aléatoire n'améliore pas l'expérience utilisateur). Aucune différence significative n'a été montrée entre les modèles. Ils contribuent également à l'amélioration de l'expérience audiovisuelle.

D.1.2 Rendu haptique pour des contenus haptique-audiovisuels basé sur un filtre perceptif

Grâce aux outils d'éditions [WRTH13, Kim13], les effets haptiques peuvent être créés indépendamment des interfaces haptiques. Par exemple des effets de mouvement comme des trajectoires ou des accélérations peuvent être définis. Les algorithmes de rendu haptique doivent donc adapter ces effets à l'interface haptique utilisée. Les limites de l'espace de travail de l'interface doivent être respectées et les transitions entre les effets doivent être gérées.

D.1.2.1 Concept

Nous proposons un nouvel algorithme de rendu haptique pour des contenus haptique-audiovisuels. Cet algorithme repose sur l'utilisation d'un filtre perceptif[1]. Ces filtres sont utilisés dans le contrôle des simulateurs de mouvements pour ramener l'appareil vers le centre de son espace de travail, sans que l'utilisateur s'en aperçoive. L'utilisation de l'espace de travail est ainsi optimisée pour que l'appareil puisse générer des effets successifs. Pour être imperceptible, ce mouvement de "remise à zéro" se fait donc sous le seuil de perception de l'utilisateur (défini par le système vestibulaire).

Dans le contexte de l'utilisation d'un bras à retour de force, le système vestibulaire n'est pas sollicité. Le filtre perceptif doit respecter les seuils de perception du système

[1] Aussi appelé *Washout Filter* en anglais

kinesthésique (perception des mouvements des membres et des forces) pour rendre des effets haptiques imperceptibles.

D.1.2.2 Implémentation

Notre algorithme a été implémenté pour le *HapSeat*. Un modèle du corps de l'utilisateur a été développé pour calculer les vitesses angulaires auxquelles sont soumises ses articulations. Ces vitesses correspondent aux seuils de perception indiquant si un mouvement est perçu ou non.

Trois profils de seuils de perception ont été définis pour exploiter différentes caractéristiques de notre algorithme. Le premier profil se base sur des valeurs psychophysiques déterminant les vitesses angulaires minimales pour ressentir un mouvements [Jon00]. Ces seuils sont bas (0.5 à $1deg.s^{-1}$) ce qui a pour conséquence de réduire l'amplitude des effets haptiques dans notre algorithme. Pour limiter ce problème deux autres profils utilisant des seuils un peu plus élevés ont été définis.

D.1.2.3 Evaluation

Ce nouvel algorithme de rendu haptique a été évalué par une étude utilisateur comptant 20 participants. Sept séquences vidéo associées à des effets haptiques de mouvement ont été présentées aux participants. Quatre rendus haptique étaient proposés: trois rendus avec un filtre perceptif (associés aux trois profils), et un rendu haptique sans filtre perceptif (condition de contrôle).

Nous avons observé que les participants préfèrent les séquences avec le filtre perceptif. Plusieurs catégories de participant se dégagent de cette étude. Certains sont sensibles à la synchronisation des effets haptiques au contenu audiovisuel alors que d'autres recherchent une certaine dynamique dans les effets. Ces résultats sont intéressants pour la conception d'effets haptiques.

Enfin notre concept a été généralisé à une séquence de dix minutes. Des effets haptiques ont été édités sur le court métrage Sintel[2]. Puis une évaluation informelle a été conduite pour comparer le rendu des effets avec et sans filtre perceptif. L'étude montre que notre concept est applicable à une "vraie" séquence haptique-audiovisuelle, et que l'algorithme de rendu améliore l'expérience utilisateur.

D.2 Partie II - Production d'effets haptiques: outils et techniques pour créer des contenus haptique-audio-visuels

Dans cette deuxième partie nous nous focalisons sur la production d'effets haptiques. Plus précisément nous proposons un nouvel outil d'édition, le *H-Studio*, et nous explorons le potentiel des retours haptiques pour créer de nouveaux effets.

[2]Crédit: Blender Fundation

D.2.1 H-Studio: un outil d'édition pour ajouter des effets haptiques et de mouvement à un contenu audiovisuel

De plus en plus de travaux proposent des systèmes de rendu haptique-audiovisuels. Mais les contenus pour ces nouvelles technologies sont encore difficiles à produire. Peu d'éditeurs existent pour faciliter cette tâche.

D.2.1.1 H-Studio

Nous proposons un nouvel outil d'édition pour créer des effets haptiques et pour les synchroniser avec un contenu audiovisuel (voir Figure D.3). Nous nous intéressons plus particulièrement à la création d'effets de mouvement.

L'innovation de cet outil repose dans la combinaison d'une interface d'édition graphique avec un bras à retour de force. De cette façon l'utilisateur peut, d'une part, créer facilement un effet de mouvement en manipulant l'appareil, et d'autre part, "previsualiser" cet effet via cet appareil. Ainsi il peut créer un effet de mouvement et avoir un aperçu du rendu haptique sans pour autant avoir à le tester sur le dispositif final, potentiellement inaccessible (salle de cinéma 4D par exemple).

Par ailleurs nous avons développé un outil de capture d'effets de mouvement, combinant une caméra et une centrale inertielle (accéléromètre, gyroscope et magnétomètre). Un contenu audiovisuel et les informations de mouvement peuvent être ainsi facilement capturés, puis importés dans le *H-Studio*. De cette façon un effet de mouvement réaliste peut être facilement conçu.

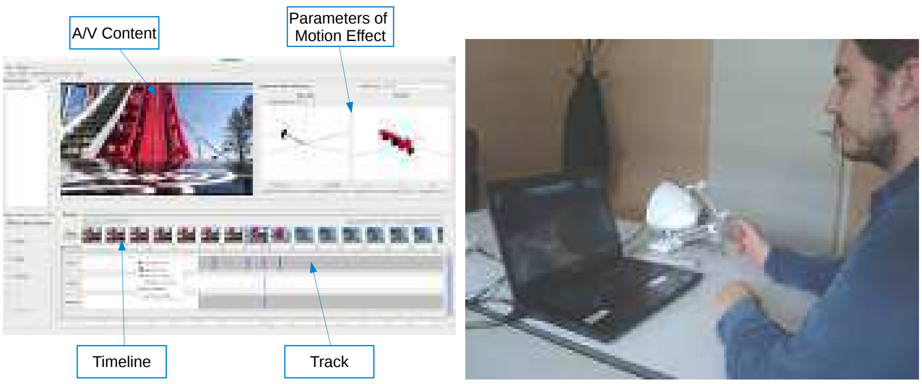


Figure D.3: Le *H-Studio*. Gauche - Capture d'écran de l'éditeur. Droite - Un utilisateur pré-visualisant un effet de mouvement.

D.2.1.2 Evaluation

Nous avons évalué le rendu des effets de mouvement capturés sur un bras à retour de force. 15 participants ont pris part à l'expérience. Il leur a été demandé de regarder 4 séquences vidéo. Les séquences étaient jouées 3 fois, chacune avec un retour haptique différent: un retour généré d'après les effets de mouvement capturés, un retour aléatoire et un retour nul. Les participants devaient évaluer les séquences haptique-audiovisuelles via un questionnaire.

Nous avons observé que les effets de mouvement capturés améliorent la qualité de l'expérience audiovisuelle (par rapport à une séquence sans effet ou avec des effets aléatoires). Par ailleurs l'utilisation d'un bras haptique suffit à donner une impression de mouvement et à enrichir le contenu vidéo.

D.2.2 Cinématographie Haptique: améliorer l'expérience audiovisuelle avec des effets haptiques basés sur des mouvements de caméra

Dans cette dernière partie de thèse nous nous sommes penchés sur l'étude de nouveaux effets haptiques pour enrichir l'expérience audiovisuelle. En effet, les retours haptiques sont principalement utilisés pour souligner des évènements physiques (type explosions, coups de feu, etc.) alors qu'ils ont le potentiel pour transmettre de l'information ou de l'émotion.

D.2.2.1 Concept

Nous proposons le concept de *Cinématographie Haptique* où les retours haptiques sont considérés comme un nouveau moyen d'expression pour les créateurs de contenus. Dans ce contexte nous introduisons une taxonomie d'effets haptiques.

Nous étudions plus précisément un type d'effet: les effets haptiques liés aux mouvements de caméra. Les mouvements de caméra ("Dutch Angle", "Vertigo", etc.) sont souvent utilisés par les réalisateurs de film pour transmettre une émotion ou une intention particulière. Notre hypothèse est qu'un effet haptique peut être utilisé pour intensifier ces mouvements de caméra.

D.2.2.2 Implémentation

Nous avons implémenté ce concept sur le *HapSeat*. Deux modèles ont été développés pour générer des effets haptiques d'après des mouvements de caméra typiques. Le premier modèle, dit modèle *Cinématique*, transpose directement les mouvements de caméra en retour haptique. L'utilisateur à l'impression de suivre la caméra. Le deuxième modèle, dit modèle *Sémantique*, génère un effet haptique qui se veut être une métaphore du mouvement de caméra. Le retour haptique a plus pour but de souligner l'intention du réalisateur (un "Dutch Angle" est utilisé pour montrer une situation instable, un "Vertigo" pour donner une sensation de vertige, etc.).

D.2.2.3 Evaluation

Notre approche a été évaluée lors d'une étude utilisateur avec 38 participants. Ces derniers devaient observer plusieurs séquences vidéo illustrant les mouvements de caméra. Pour chaque séquence, 4 retours haptiques étaient testés: un généré d'après le modèle *Cinématique*, un d'après le modèle *Sémantique*, un aléatoire et un nul.

Nous avons observé que ces nouveaux effets haptiques améliorent l'expérience audiovisuelle. Le modèle *Cinématique* semble convenir dans tous les cas de figure, et les métaphores du modèle *Sémantique* sont appréciées dans certaines conditions. La dynamique du retour haptique doit être similaire à celle de la vidéo pour ne pas générer un sentiment d'incohérence chez l'utilisateur.

D.3 Conclusion

Dans ce manuscrit de thèse nous avons étudié le potentiel des retours haptiques pour améliorer l'expérience audiovisuelle. Nous avons suivi deux axes de recherche. Le premier axe correspond au rendu d'effets haptiques pour des contenus audiovisuels. Le deuxième axe correspond à la production de contenus haptique-audiovisuels. La thèse est découpée en deux parties suivant ces deux axes.

Plus précisément, dans la première partie nous avons proposé un nouvel appareil pour générer des sensations de mouvement lors du visionnage d'une séquence vidéo. Trois bras à retour de force sont utilisés pour stimuler les bras et la tête de l'utilisateur. Une étude utilisateur a montré que ce dispositif enrichit l'expérience audiovisuelle et génère une sensation de mouvement réaliste. Dans un deuxième temps nous nous sommes penchés sur les algorithmes de rendu haptiques. Ceux-ci ne sont pas adaptés à une utilisation dans un contexte audiovisuel où les effets haptiques peuvent être conçus indépendamment de l'appareil utilisé. Nous avons donc proposé un nouvel algorithme de rendu intégrant un filtre perceptif. Une nouvelle étude utilisateur a été conduite et valide notre approche.

La deuxième partie de la thèse s'est focalisée sur la production d'effets haptiques. Un nouvel outil d'édition a été développé, facilitant la création d'effets de mouvement. Cet outil couple un appareil à retour de force, un appareil de capture d'effets, et une interface graphique. Ainsi il est possible de créer facilement des effets de mouvement et de les "pré-visualiser". Une étude utilisateur a montré l'intérêt d'un tel système. Enfin nous avons exploré les combinaisons haptique-audiovisuelles dans le but de proposer de nouveaux effets haptiques. Plus particulièrement nous avons proposé des effets haptiques basés sur des effets de mouvement de caméra. Une étude utilisateur a aussi montré la pertinence de notre concept.

Des travaux de recherche sont encore nécessaires pour faire en sorte que l'haptique-audiovisuel devienne une technologie mature. Mais les récentes contributions scientifiques et les développements technologiques montrent l'intérêt grandissant envers ce jeune champ d'étude. Il ne fait aucun doute que les technologies haptiques ont un fort potentiel pour enrichir l'expérience audiovisuelle, et qu'elles seront de plus en plus utilisées pour créer des applications immersives. Nous espérons que les travaux présentés dans ce manuscrit contribueront à leur développement.

Author's publications

Articles

Journal papers

J1. **F. Danieau**, A. Lécuyer, P. Guillotel, J. Fleureau, N. Mollet and M. Christie. "Enhancing audiovisual experience with haptic feedback: a survey on HAV". *IEEE Transactions on Haptics*, vol. 6, no. 2, pp. 193–205, 2013.

J2. **F. Danieau**, A. Lécuyer, P. Guillotel, J. Fleureau, N. Mollet and M. Christie. "Toward Haptic Cinematography: Enhancing Movie Experience with Haptic Effects based on Cinematographic Camera Motions", *ACM Multimedia*, In Press.

Conference papers

C1. **F. Danieau**, J. Fleureau, A. Cabec, P. Kerbiriou, P. Guillotel, N. Mollet, M. Christie and A. Lécuyer. "A Framework for Enhancing Video Viewing Experience with Haptic Effects of Motion". *IEEE Haptics Symposium*, pp. 541–546, 2012.

C2. **F. Danieau**, J. Fleureau, P. Guillotel, N. Mollet, M. Christie and A. Lécuyer. "HapSeat: Producing Motion Sensation with Multiple Force-feedback Devices Embedded in a Seat". *ACM VRST*, pp. 69–76, 2012.

Demonstrations

D1. **F. Danieau**, J. Bernon, J. Fleureau, P. Guillotel, N. Mollet, M. Christie and A. Lécuyer. "H-Studio: An Authoring Tool for Adding Haptic and Motion Effects to Audiovisual Content". *ACM UIST*, pp. 83–84, 2013.

D2. **F. Danieau**, J. Fleureau, P. Guillotel, N. Mollet, M. Christie and A. Lécuyer. "HapSeat: a novel approach to simulate motion in audiovisual experiences". *ACM SIGGRAPH Emerging Technologies*, 2013.

D2. **F. Danieau**, J. Fleureau, P. Guillotel, N. Mollet, M. Christie and A. Lécuyer. "HapSeat: a novel approach to simulate motion in a consumer environment". *ACM CHI Extended Abstracts on Human Factors in Computing Systems*, pp. 3035–3038, 2013.

Submitted

S1. **F. Danieau**, A. Lécuyer, P. Guillotel, J. Fleureau, N. Mollet and M. Christie. “Haptic Rendering for HAV based on a Washout Filter”. Submitted to *IEEE Transactions on Haptics.*

Patents

P1. J. Fleureau, **F. Danieau**, P. Guillotel, and A. Lécuyer. “Methods to command a haptic renderer from real motion data.” WO2013041152.

P2. **F. Danieau**, F. Fleureau, P. Guillotel, N. Mollet, A. Lécuyer and M. Christie. “Haptic chair for motion simulation”. WO2013153137.

P3. **F. Danieau**, F. Fleureau, P. Guillotel, N. Mollet and A. Lécuyer. “Method to render global 6 DoF motion effect with multiple local force-feedback”. WO2013153086.

Bibliography

[AMB] Ambx website. http://http://www.ambx.com. Accessed: 2013-11-01.

[AMS11] Jason Alexander, Mark T. Marshall, and Sriram Subramanian. Adding haptic feedback to mobile tv. In *ACM Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1975–1980, 2011.

[Ber00] A. Berthoz. *The brain's sense of movement*. Harward University Press, 2000.

[BMO+02] Andy Brady, B. MacDonald, Ian Oakley, Stephen Hughes, and S. O'Modhrain. Relay: a futuristic interface for remote driving. In *Eurohaptics*, 2002.

[BS02] S. James Biggs and M.A. Srinivasan. Haptic interfaces. *Handbook of virtual Environments*, pages 93 – 116, 2002.

[BUT] Buttkicker website. http://www.thebuttkicker.com. Accessed: 2013-11-01.

[CCO+08] Dongju Chi, Donghyun Cho, Sungjin Oh, Kyunngkoo Jun, Yonghee You, Hwanmun Lee, and Meeyoung Sung. Sound-specific vibration interface using digital signal processing. In *IEEE Computer Science and Software Engineering*, volume 4, pages 114–117, 2008.

[CES09] Jongeun Cha, Mohamad Eid, and Abdulmotaleb El Saddik. Touchable 3D video system. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 5(4):1–25, October 2009.

[CHK+09] Jongeun Cha, Yo-Sung Ho, Yeongmi Kim, Jeha Ryu, and Ian Oakley. A framework for haptic broadcasting. *IEEE Multimedia*, 16(3):16–27, July 2009.

[CJ4] Cj 4dxplex website. http://cj4dx.com/. Accessed: 2013-11-01.

[CLM08] Christel Chamaret and Olivier Le Meur. Attention-based video reframing: validation using eye-tracking. In *IEEE 19th International Conference on Pattern Recognition*, pages 1–4, 2008.

[CLY11] Bumsuk Choi, Eun-Seo Lee, and Kyoungro Yoon. Streaming media with sensory effect. In *IEEE International Conference on Information Science and Applications*, pages 1–6. IEEE, 2011.

[CO05] Angela Chang and Conor O'Sullivan. Audio-haptic feedback in mobile phones. *ACM Extended Abstracts of the Computer Human Interaction Conference on Human factors in Computing Systems*, pages 1264–1267, 2005.

[CON08] Marc Christie, Patrick Olivier, and Jean-Marie Normand. Camera Control in Computer Graphics. *Computer Graphics Forum*, 27(8):2197–2218, December 2008.

[CSKR07] Jongeun Cha, Yongwon Seo, Yeongmi Kim, and Jeha Ryu. An Authoring/Editing Framework for Haptic Broadcasting: Passive Haptic Interactions using MPEG-4 BIFS. *Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)*, pages 274–279, 2007.

[CSP02] S.F. Chang, Thomas Sikora, and A. Purl. Overview of the MPEG-7 standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):688–695, 2002.

[Das00] B. Dasgupta. The Stewart platform manipulator: a review. *Mechanism and Machine Theory*, 35(1):15–40, 2000.

[DBO] D-box website. http://www.d-box.com. Accessed: 2013-11-01.

[Dio97] José Dionisio. Projects in VR virtual hell: a trip through the flames. *IEEE Computer Graphics And Applications*, pages 11–14, 1997.

[DSP91] Glorianna Davenport, Thomas Aguierre Smith, and Natalio Pincever. Cinematic Primitives for Multimedia. *IEEE Computer Graphics and Applications*, 11(4):67–74, 1991.

[DWAD10] E.O. Dijk, A. Weffers-Albu, and T. De Zeeuw. A tactile actuation blanket to intensify movie experiences with personalised tactile effects. In *International Conference on Intelligent Technologies for Interactive Entertainment*, pages 2–3, 2010.

[EAAE08] Mohamad Eid, Sheldon Andrews, Atif Alamri, and A. El Saddik. HAMLAT: a HAML-based authoring tool for haptic application development. *Haptics: Perception, Devices and Scenarios*, pages 857–866, 2008.

[EfEOS06] Fayez El-far, Mohamad Eid, Mauricio Orozco, and Abdulmotaleb Saddik. Haptic Applications Meta-Language. *IEEE International Symposium on Distributed Simulation and Real-Time Applications*, pages 261–264, 2006.

[EM03] M.J. Enriquez and K.E. MacLean. The hapticon editor: a tool in support of haptic communication research. *Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 356–362, 2003.

[EOEC11] Abdulmotaleb El Saddik, Mauricio Orozco, Mohamad Eid, and Jongeun Cha. *Haptic technologies - bringing Touch to multimedia*. Springer, 2011.

[FGHT12] J. Fleureau, P. Guillotel, and Q. Huynh-Thu. Physiological-based affect event detector for entertainment video applications. *IEEE Transactions on Affective Computing*, 2012.

[FLA] Adobe flash website. http://www.adobe.com/products/flash.html. Accessed: 2013-11-01.

[GALSC12] Frédéric Giraud, Michel Amberg, Betty Lemaire-Semail, and Géry Casiez. Design of a transparent tactile stimulator. In *Haptics Symposium (HAPTICS)*, pages 485–489. IEEE, 2012.

[GMS06] D. Gaw, D. Morris, and K. Salisbury. Haptically annotated movies: reaching out and touching the silver screen. *Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 287–288, 2006.

[HACH+04] Vincent Hayward, Oliver R. Astley, Manuel Cruz-Hernandez, Danny Grant, and Gabriel Robles-De-La-Torre. Haptic interfaces and devices. *Sensor Review*, 24(1):16–29, 2004.

[Hei62] M.L. Heilig. Sensorama simulator, August 1962. U.S. Patent 3050870.

[HESG08] Abdelwahab Hamam, Mohamad Eid, Abdulmotaleb El Saddik, and Nicolas D. Georganas. A quality of experience model for haptic user interfaces. In *Proceedings of the 2008 Ambi-Sys workshop on Haptic user interfaces in ambient media systems*, pages 1–6, 2008.

[HGE10] Abdelwahab Hamam, N.D. Georganas, and A. El Saddik. Effect of haptics on the quality of experience. In *IEEE International Symposium on Haptic Audio-Visual Environments and Games (HAVE)*, pages 1–6, 2010.

[HJZ+02] L. R. Harris, M. R. Jenkin, D. Zikovitz, F. Redlick, P. Jaekl, U. T. Jasiobedzka, H. L. Jenkin, and R. S. Allison. Simulating self-motion I: cues for the perception of motion. *Virtual Reality*, 6(2):75–85, September 2002.

[HS04] K.S. Hale and K.M. Stanney. Deriving haptic design guidelines from human physiological, psychophysical, and neurological foundations. *IEEE Computer Graphics and Applications*, 24(2):33–39, 2004.

[HS06] Peter F. Hokayem and Mark W. Spong. Bilateral teleoperation: an historical survey. *Automatica*, 42(12):2035–2057, 2006.

[HTBLC11] Quan Huynh-Thu, Marcus Barkowsky, and Patrick Le Callet. The importance of visual attention in improving the 3d-tv viewing experience: overview and new perspectives. *IEEE Transactions on Broadcasting*, 57(2):421–431, 2011.

[HTIS10] Takayuki Hoshi, Masafumi Takahashi, Takayuki Iwamoto, and Hiroyuki Shinoda. Noncontact tactile display based on radiation pressure of airborne ultrasound. *IEEE Transactions on Haptics*, 3(3):155–165, 2010.

[HTWM04] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 34(3):334–352, 2004.

[IMM] Immersion website. http://www.immersion.com/products/integrator/. Accessed: 2013-11-01.

[IP11] Ali Israr and Ivan Poupyrev. Tactile brush: drawing on skin with a tactile grid display. *ACM CHI*, pages 2019–2028, 2011.

[ISO05] ISO/IEC 14496-11. Information technology – coding of audio-visual objects – Part 11: scene description and application engine, 2005.

[Jai04] Ramesh Jain. Quality of experience. *IEEE Multimedia*, 65(1):1–2, 2004.

[JB02] Lynette A. Jones and Michal Berris. The psychophysics of temperature perception and thermal-interface design. In *IEEE Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 137–142, 2002.

[JMF99] A K Jain, M N Murty, and P J Flynn. Data Clustering: A Review. *ACM computing surveys (CSUR)*, 31(3):264–323, 1999.

[Jon00] Lynette A Jones. Kinesthetic sensing. *Human and Machine Haptics*, pages 1–10, March 2000.

[KCRO10] Y Kim, J Cha, Jeha Ryu, and Ian Oakley. A tactile glove design and authoring system for immersive multimedia. *IEEE Multimedia*, 17(3):34–45, 2010.

[KFN05] K.J. Kuchenbecker, J. Fiene, and G. Niemeyer. Event-based haptics and acceleration matching: portraying and assessing the realism of contact. *Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 381–387, 2005.

[Kil08] K. Kilkki. Quality of experience in communications ecosystem. *Journal of universal computer science*, 14(5):615–624, 2008.

[Kim13] Sang-Kyun Kim. Authoring multisensorial content. *Signal Processing: Image Communication*, 28(2):162–167, February 2013.

[KIP13] Seung-Chan Kim, Ali Israr, and Ivan Poupyrev. Tactile rendering of 3d features on touch surfaces. In *ACM UIST*, pages 531–538, 2013.

[KLC12] Myongchan Kim, Sungkil Lee, and Seungmoon Choi. Saliency-driven tactile effect authoring for real-time visuotactile feedback. In *EuroHaptics*, pages 258–269. Springer, 2012.

[KPK+11] Y. Kim, Sunyoung Park, Hyungon Kim, Hyerin Jeong, and Jeha Ryu. Effects of different haptic modalities on students' understanding of physical phenomena. In *IEEE World Haptics Conference*, pages 379–384, 2011.

[KSH09] Youngjae Kim, Heesook Shin, and Minsoo Hahn. A bidirectional haptic communication framework and an authoring tool for an instant messenger. In *IEEE Advanced Communication Technology*, volume 3, pages 2050–2053, 2009.

[LC13] Jaebong Lee and Seungmoon Choi. Real-Time Perception-Level Translation from Audio Signals to Vibrotactile Effects. In *ACM CHI*, pages 2567–2576, 2013.

[LCB+09] Paul Lemmens, Floris Crompvoets, Dirk Brokken, Jack Van Den Eerenbeemd, and Gert-jan De Vries. A body – conforming tactile jacket to enrich movie viewing. *Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 7–12, 2009.

[LD03] S. D. Laycock and a. M. Day. Recent developments and applications of haptic devices. *Computer Graphics Forum*, 22(2):117–132, 2003.

[LLC+05] B.C. Lee, Junhun Lee, Jongeun Cha, Changhoon Seo, and Jeha Ryu. Immersive live sports experience with vibrotactile sensation. *Human-Computer Interaction - INTERACT*, pages 1042–1045, 2005.

[LVJ+04] Anatole Lécuyer, Manuel Vidal, Olivier Joly, Christine Mégard, and Alain Berthoz. Can haptic feedback improve the perception of self-motion in virtual reality? In *Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 208–215. IEEE, 2004.

[LYNH06] R.W. Lindeman, Yasuyuki Yanagida, Haruo Noma, and Kenichi Hosaka. Wearable vibrotactile systems for virtual contact and information display. *Virtual Reality*, 9(2):203–213, 2006.

[Mac96] K MacLean. The 'haptic camera': A technique for characterizing and playing back haptic properties of real environments. In *Proceedings of ASME dynamic systems and control division*, volume 58, pages 459–467, 1996.

[Mas98] J. V. Mascelli. *The five C's of the Cinematography: Motion picture filming techniques*. Silman-James Press, 1998.

[MED] Mediamation website. http://mediamation.com/. Accessed: 2013-11-01.

[MIC06] Regan L Mandryk, Kori M Inkpen, and Thomas W Calvert. Using psychophysiological techniques to measure user experience with entertainment technologies. *Behavior & Information Technology*, 25(2):141–158, 2006.

[MIWB02] Michael Meehan, Brent Insko, Mary Whitton, and Frederick P. Jr. Brooks. Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics*, 21(3):645–652, 2002.

[MKT+05] Troy L. McDaniel, Kanav Kahol, Priyamvada Tripathi, David P. Smith, Laura Bratton, Ravi Atreya, and Sethuraman Panchanathan. A methodology to establish ground truth for computer vision algorithms to estimate haptic features from visual images. *IEEE International Workshop on Haptic Audio Visual Environments and their Applications*, pages 94–99, 2005.

[MPE11] MPEG Group. ISO/IEC FDIS 23005-3 Information technology – media context and control – Part 3: sensory information. 2011.

[MTB06] N. Magnenat-Thalmann and U. Bonanni. Haptics in virtual reality and multimedia. *IEEE Multimedia*, 13(3):6–11, 2006.

[NOV] Novint website. http://www.novint.com. Accessed: 2013-11-01.

[NTWO09] S. Nanayakkara, Elizabeth Taylor, L. Wyse, and S.H. Ong. An enhanced musical experience for the deaf: design and evaluation of a music display and a haptic chair. In *ACM International conference on Human factors in computing systems*, pages 337–346, 2009.

[OHA+08] Sayaka Ooshima, Yuki Hashimoto, Hideyuki Ando, Junji Watanabe, and Hiroyuki Kajimoto. Simultaneous presentation of tactile and auditory motion to the abdomen to present the feeling of being slashed. *IEEE SICE Annual Conference*, pages 467–471, 2008.

[OLB09] N. Ouarti, A. Lécuyer, and A. Berthoz. Method for simulating specific movements by haptic feedback, and device implementing the method, September 2009. French Patent $N°$ 09 56406.

[OO03] S. O'Modhrain and Ian Oakley. Touch TV: adding feeling to broadcast media. In *European Conference on Interactive Television: from Viewers to Actors*, 2003.

[OO04] S. O'Modhrain and I. Oakley. Adding interactivity: active touch in broadcast media. In *IEEE Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 293–294, 2004.

[OSS13] Marianna Obrist, Sue Ann Seah, and Sriram Subramanian. Talking about tactile experiences. In *ACM CHI*, pages 1659–1668, 2013.

[PHA] Geomagic website. http://http://geomagic.com/en/products/phantom-premium/overview. Accessed: 2013-11-01.

[PWN+] Saurabh Palan, Ruoyao Wang, Nathaniel Naukam, Li Edward, and Katherine J. Kuchenbecker. Tactile Gaming Vest (TGV). http://iroboticist.com/2010/03/26/tgv/. Accessed: 2013-09-30. Hands-on demonstration presented at IEEE Haptics Symposium 2010, Boston, Massachusetts.

[RAC10] M Abdur Rahman, Abdulmajeed Alkhaldi, and Jongeun Cha. Adding haptic feature to YouTube. *International conference on Multimedia*, pages 1643–1646, 2010.

[RC08] Jonghyun Ryu and Seungmoon Choi. posVibEditor: graphical authoring tool of vibrotactile patterns. *IEEE International Workshop on Haptic Audio visual Environments and Games*, (October):120–125, October 2008.

[Rei04] M. Reiner. The role of haptics in immersive telecommunication environments. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(3):392–401, March 2004.

[RS11] Shahzad Rasool and Alexei Sourin. Haptic interaction with 2D images. In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*, volume 1, pages 13–22, 2011.

[RSP] B.E. Riecke and J. Schulte-Pelkum. Using the perceptually oriented approach to optimize spatial presence & ego-motion simulation. Max Planck Institute for Biological Cybernetics, Technical Report *N*°153, 2006.

[RSPCB05] Bernhard E. Riecke, Jörg Schulte-Pelkum, Franck Caniard, and Heinrich H. Bülthoff. Towards lean and elegant self-motion simulation in virtual reality. In *Virtual Reality*, pages 131–138, 2005.

[Sab06] A. M. Sabatini. Quaternion-based extended kalman filter for determining orientation by inertial and magnetic sensing. *IEEE Transactions on Biomedical Engineering*, 53(7):1346–1356, 2006.

[SACH10] Suchul Shin, Yangwoo Ahn, Jaejung Choi, and Soonhung Han. A Framework for Producing Motion Effects of 4D Films using Washout Filters. In *Asian Conference on Design and Digital Engineering*, pages 758–765, 2010.

[SCB04] Kenneth Salisbury, Francois Conti, and Federico Barbagli. Haptic rendering: introductory concepts. *IEEE computer graphics and applications*, 24(2):24–32, 2004.

[Shi93] K.B. Shimoga. A survey of perceptual feedback issues in dexterous telemanipulation. II. Finger touch feedback. In *IEEE Virtual Reality Annual International Symposium*, pages 271–279, 1993.

[SKI05] Yuriko Suzuki, Minoru Kobayashi, and Air Jet Interface. Air jet driven force feedback in virtual reality. *IEEE computer graphics and applications*, 25(1):44–7, 2005.

[SLK+07] Yongwon Seo, Beom-chan Lee, Yeongmi Kim, Jong-phil Kim, and Jeha Ryu. K-HapticModeler: a haptic modeling scope and basic framework. *International Workshop on Haptic Audio Visual Environments and their Applications*, pages 136 – 141, 2007.

[SM07] Jocelyn Smith and K. MacLean. Communicating emotion through a haptic link: design space and methodology. *International Journal of Human-Computer Studies*, 65(4):376–387, 2007.

[SPGI13] Rajinder Sodhi, Ivan Poupyrev, Matthew Glisson, and Ali Israr. AIREAL: interactive tactile experiences in free air. *ACM SIGGRAPH*, 32(4), 2013.

[SSS+10] Michael Schuenke, Erik Schulte, Udo Schumacher, Lawrence M. Ross, Edward D. Lamperti, and Markus Voll. *General Anatomy and Musculoskeletal System*. Thieme Medical Publishers Inc., 2010.

[STMA10] A. Shah, S. Teuscher, E. McClain, and J. Abbott. How to build an inexpensive 5-dof haptic device using two novint falcons. In *Haptics: Generating and Perceiving Tangible Sensations*, pages 136–143, 2010.

[SVKV01] M.J. Schuemie, P. Van Der Straaten, M. Krijn, and C.A.P.G. Van Der Mast. Research on presence in virtual reality: a survey. *CyberPsychology & Behavior*, 4(2):183–201, 2001.

[SW08] Alexei Sourin and Lei Wei. Visual immersive haptic rendering on the web. *ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, page 1, 2008.

[TA08] T. Tullis and W. Albert. *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Morgan Kaufmann, 2008.

[TB09] R. Thompson and C. Bowen. *Grammar of the Edit - Second Edition*. Focal Press, 2009.

[TBN+04] M. Trousselard, P-A. Barraud, V. Nougier, C. Raphel, and C. Cian. Contribution of tactile and interoceptive cues to the perception of the direction of gravity. *Cognitive brain research*, 20(3):355–62, 2004.

[Tho06] G A Thomas. Sports Real-Time Camera Pose Estimation for Augmenting Scenes. In *CVMP*, pages 10–19, 2006.

[TM07] Tinne Tuytelaars and Krystian Mikolajczyk. Local invariant feature detectors: a survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3):177–280, 2007.

[TNG] Tngames website. http://tngames.com. Accessed: 2013-11-01.

[TNP+09] D. Tsetserukou, A. Neviarouskaya, H. Prendinger, N. Kawakami, and S. Tachi. Affective haptics in emotional communication. In *IEEE International Conference on Affective Computing and Intelligent Interaction and Workshops*, number 1, pages 1–6, 2009.

[TV07] Ba Tu Truong and Svetha Venkatesh. Video abstraction: a systematic review and classification. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 3(1), 2007.

[uR08] S. ur Rehman. Turn your mobile into the ball: rendering live football game using vibration. *IEEE Transactions on Multimedia*, 10(6):1022–1033, October 2008.

[vE02] Jan B.F. van Erp. Guidelines for the use of vibro-tactile displays in human computer interaction. In *Eurohaptics*, volume 2002, pages 18–22, 2002.

[VLVK05] A. Valjamae, P. Larsson, D. Vastfjall, and M. Kleiner. Travelling without moving: auditory scene cues for translational self-motion. *Proceeding of International Conference of Auditory Display*, 2005.

[Wel93] Chris Welman. Inverse kinematics and geometric constraints for articulated figure manipulation. Master's thesis, Simon Fraser University, 1993.

[WHZ+01] Dapeng Wu, Yiwei Thomas Hou, Wenwu Zhu, Ya-Qin Zhang, and Jon M. Peha. Streaming video over the Internet: approaches and directions. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(3):282–300, 2001.

[WRTH12] Markus Waltl, Benjamin Rainer, Christian Timmerer, and Hermann Hellwagner. A toolset for the authoring, simulation, and rendering of sensory experiences. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 1469–1472, 2012.

[WRTH13] Markus Waltl, Benjamin Rainer, Christian Timmerer, and Hermann Hellwagner. An end-to-end tool chain for Sensory Experience based on MPEG-V. *Signal Processing : Image Communication*, 28(2):136–150, 2013.

[WS98] Bob G. Witmer and Michael J. Singer. Measuring presence in virtual environments : A presence questionnaire. *Presence*, pages 225–240, 1998.

[WT10] M Waltl and C Timmerer. Increasing the user experience of multimedia presentations with sensory effects. In *11th International Workshop on*

Image Analysis for Multimedia Interactive Services (WIAMIS), pages 1–4, 2010.

[YAMS06] Takehiko Yamaguchi, Ayumu Akabane, Jun Murayama, and Makoto Sato. Automatic generation of haptic effect into published 2D graphics. In *Eurohaptics*, 2006.

[YCLL10] Kyoungro Yoon, Bumsuk Choi, E.S. Lee, and T.B. Lim. 4-D broadcasting with MPEG-V. In *IEEE Multimedia Signal Processing*, number 2, pages 257–262, 2010.

[Yoo13] Kyoungro Yoon. End-to-end framework for 4-D broadcasting based on MPEG-V standard. *Signal Processing: Image Communication*, 28(2):127–135, February 2013.

[ZCPR03] W Zhao, R. Chellappa, PJ Phillips, and A Rosenfeld. Face recognition: a literature survey. *ACM Computing Surveys (CSUR)*, 35(4):399–458, 2003.

[ZK01] Tong Zhang and C.-C.J. Kuo. Audio content analysis for online audiovisual data segmentation and classification. *IEEE Transactions on Speech and Audio Processing*, 9(4):441–457, 2001.

[ZWE+05] Xingquan Zhu, Xindong Wu, Ahmed K. Elmagarmid, Zhe Feng, and Lide Wu. Video data mining: semantic indexing and event detection from the association perspective. *IEEE Transactions on Knowledge and Data Engineering*, 17(5):665–677, 2005.

List of Figures

List of Tables

Résumé

Les technologies haptiques, stimulant le sens du toucher, sont utilisées depuis des années dans des applications de réalité virtuelle et de téléopération pour accroître l'immersion de l'utilisateur. Elles sont en revanche très peu employées dans les systèmes audiovisuels comme les cinémas. L'objectif de cette thèse est d'exploiter le potentiel des retours haptiques pour les contenus audiovisuels.

Dans la première partie de la thèse, nous nous intéressons au rendu d'effets haptiques lors du visionnage d'une vidéo. Nous présentons tout d'abord un appareil générant des sensations de mouvements à 6 degrés de liberté. Au lieu de mettre tout le corps de l'utilisateur en mouvement, comme cela est fait avec les simulateurs de mouvements traditionnels, seulement la tête et les mains sont stimulées. Ce dispositif permet ainsi d'enrichir l'expérience audiovisuelle. Nous nous intéressons ensuite aux algorithmes de rendu d'effets haptiques dans un contexte audiovisuel. La combinaison de retours haptiques et de séquences vidéo amène de nouveaux problèmes lors du rendu haptique. Nous proposons un nouvel algorithme adapté à ce contexte.

Dans la seconde partie de la thèse, nous nous concentrons sur la production d'effets haptiques. Premièrement nous présentons un nouvel outil d'édition graphique. Celui-ci propose trois méthodes d'interaction pour créer des effets de mouvement et pour les synchroniser avec une vidéo. De plus, cet outil permet de ressentir les effets créés. Ensuite nous nous penchons sur les combinaisons haptiques et audiovisuelles. Dans une nouvelle approche nommée *Cinématographie Haptique*, nous explorons le potentiel des effets haptiques pour créer de nouveaux effets dédiés aux réalisateurs de films.

Abstract

Haptic technology, stimulating the sense of touch, is used for years in virtual reality and teleoperation applications for enhancing the user immersion. Yet it is still underused in audiovisual systems such as movie theaters. The objective of this thesis is thus to exploit the potential of haptics for audiovisual content.

In the first part of this Ph.D. thesis, we address the haptic rendering in video viewing context. We first present a new device providing 6 degrees of freedom motion effects. Instead of moving the whole user's body, as it is traditionally done with motion platform, only the head and hands are stimulated. This device allows thus to enrich the audiovisual experience. Then we focus on the haptic rendering of haptic-audiovisuals. The combination of haptic effects and video sequences yields new challenges for the haptic rendering. We introduce a new haptic rendering algorithm to tackle these issues.

The second part of this Ph.D. is dedicated to the production of haptic effects. We first present of novel authoring tool. Three editing methods are proposed to create motion effects and to synchronize them to a video. Besides, the tool allows to preview motion effects thanks to a force-feedback device. Then we study combinations of haptic feedback and audiovisual content. In a new approach, the *Haptic Cinematography*, we explore the potential of haptic effects to create new effects dedicated to movie makers.